

Marcos Daniel Valadão Baroni

Algoritmo Híbrido *Intelligent Water Drops*
para o Problema da Mochila Multidimensional

Projeto para Curso de Doutorado em Ciência da
Computação

Vitória, 28 de maio de 2013

Programa de Pós-Graduação em Informática
Centro Tecnológico – Universidade Federal do Espírito Santo
Av. Fernando Ferrari, s/n, Vitória - ES

Sumário

1	Introdução	1
2	Objetivos	3
3	Justificativa	3
4	Metodologia	7
5	Cronograma	7

1 Introdução

A otimização combinatória é o ramo da computação que trata de problemas cujo objetivo é encontrar um elemento ótimo dentre um conjunto finito de elementos [Schrijver, 2003]. Nos últimos anos o estudo da otimização combinatória tem crescido devido ao aparecimento de complexos problemas deste caráter em várias áreas, como produção industrial e economia.

O *problema da mochila multidimensional* (ou, em inglês, *multidimensional knapsack problem* ou *MKP*) é um problema de otimização combinatória que pode ser aplicado a vários problemas práticos, como por exemplo, carregamento de carga [Shih, 1979], corte de estoque [Gilmore and Gomory, 1966], seleção de projetos [Lu et al., 1999] e alocação de processadores em sistemas distribuídos [Gavish and Pirkul, 1982].

Por ser de grande relevância e difícil resolução, o MKP tem sido bastante estudado. Este, pode ser definido pelo problema de otimização seguinte [Puchinger et al., 2010]:

$$\text{maximizar } z = \sum_{j=1}^n p_j x_j \quad (1)$$

$$\text{sujeito a } \sum_{j=1}^n w_{ij} x_j \leq c_i, \quad i = 1, \dots, m, \quad (2)$$

$$x_j \in \{0, 1\}, \quad j = 1, \dots, n. \quad (3)$$

O problema define n elementos e m recursos, cada um com capacidade $c_i > 0$. Para a seleção de um elemento j há um lucro p_j , mas também um gasto w_{ij} do recurso i . O objetivo é selecionar um subconjunto de elementos que totalize o maior lucro, não ultrapassando as capacidades dos recursos.

No caso em que $m = 1$ o problema é chamado simplesmente de *problema da mochila* (knapsack problem). Sabe-se que o problema está na classe *NP-completo*, por ser um dos 21 problemas NP-completos apresentados por Richard Karp em 1972 [Karp, 1972, Garey and Johnson, 1979].

Com o crescimento em larga escala dos problemas de otimização combinatória, métodos simples utilizados até então para a sua resolução tornaram-se

inviáveis, devido à complexidade, exigindo assim o desenvolvimento de métodos mais eficientes.

A utilização de métodos exatos para problemas NP-completos garantem otimalidade da solução, porém tornam-se inviáveis, na maioria das vezes, mesmo para instâncias de tamanho médio, devido ao tempo exponencial necessário para a execução dos métodos. Algoritmos de aproximação são capazes de dar garantia na qualidade da solução, demandando um menor tempo de execução, porém, são geralmente algoritmos difíceis de serem formulados e a garantia de solução dada por vezes não é satisfatória.

Por estes motivos cada vez mais têm-se recorrido às *heurísticas*: métodos não exatos que demandam um menor tempo de execução. Apesar de não proverem garantia na qualidade de solução, a utilização de heurísticas, após os devidos ajustes, tem apresentado em geral resultados satisfatórios.

Além das heurísticas desenvolvidas especificamente para um determinado problema existem ainda as chamadas *metaheurísticas* que são heurísticas com um caráter genérico. Este caráter genérico possibilita a sua aplicação em vários tipos de problemas, geralmente através de poucas adaptações [Dorigo, 2013]. Podem possuir ainda variações em suas aplicações ou mesmo serem aplicadas em conjunto com uma outra heurística, gerando um método híbrido, mais efetivo por aliarem as qualidades presentes em cada uma delas [Blum and Roli, 2003].

Recentemente foi proposta por Hamed Shah-Hosseini uma metaheurística chamada *intelligent water drops (IWD)* que baseia-se no comportamento das águas ao fluírem através de um rio [Shah-Hosseini, 2007]. Cada porção de água percorre um caminho no espaço que depende, não só do ambiente ao redor, mas também do tamanho e velocidade da porção d'água. Ao passar por um caminho, a porção de água modifica o ambiente, podendo ganhar velocidade e tamanho ao carregarem um pouco do solo. Tem-se no final do processo um caminho traçado da nascente ao ponto em que o rio desagua.

A proposta desta nova metaheurística foi seguida pela sua aplicação ao MKP, feita pelo mesmo autor [Shah-Hosseini, 2008], cujo experimento alcançou soluções bem próximas ao ótimo e bem competitivas quando comparadas à outras metaheurísticas aplicadas ao MKP por trabalhos conhecidos da literatura.

Devido a sua importância e fácil modelagem o MKP têm sido utilizado ultimamente como um *problema teste* na avaliação de novas metaheurísticas. O objetivo principal desta proposta é desenvolver uma heurística híbrida para a resolução do MKP, aliando a metaheurística IWD a um *algoritmo de estimação de distribuição*.

Um algoritmo de estimação de distribuição (AED) é um algoritmo probabilístico que propõe encontrar boas soluções para problemas de otimização, com o auxílio de um modelo probabilístico, proposto a partir de uma amostra de soluções. Detalhes da utilização do AED à metaheurística IWD são apresentados nas próximas seções.

Também pretende-se aplicar a heurística desenvolvida aqui a um problema real da indústria de distribuição de energia elétrica brasileira.

Na Seção seguinte apresentaremos os objetivos deste trabalho. Na Seção 3 apresentaremos a justificativa para a pesquisa. Na Seção 4 trataremos da

metodologia que será utilizada no trabalho. Na Seção 5 apresentaremos o cronograma do projeto.

2 Objetivos

Esta proposta tem como objetivo o aperfeiçoamento da metaheurística IWD aplicado ao MKP, através da utilização de um algoritmo de estimação de distribuição, gerando uma heurística híbrida para o problema em questão. A heurística será implementada e ajustes e melhorias serão feitas. Testes computacionais seguirão a implementação com o objetivo de comparar sua eficácia na resolução do MKP à outros métodos já existentes.

Os objetivos específicos do projeto são os seguintes:

- Realizar ampla revisão bibliográfica sobre o problema tratado (MKP) e os algoritmos envolvidos (IWD e AED);
- Implementar a heurística híbrida à resolução do problema, realizando os devidos ajustes;
- Propor melhorias à heurística implementada;
- Realizar comparações entre a heurística implementada e os métodos de resolução já existentes através de testes computacionais;
- Aplicar a heurística a um problema real da indústria de distribuição de energia elétrica brasileira e comparar com os resultados obtidos pelo método já existente.

3 Justificativa

O MKP também pode ser interpretado como uma formulação geral para problemas de programação inteira 0-1 com coeficientes não negativos, como é possível concluir observando sua definição. Dessa forma o MKP pode ser resolvido utilizando-se os métodos exatos já conhecidos para tratar o problema de programação inteira 0-1, como o *método aditivo de Balas* [Balas, 1965]. Porém, devido a complexidade computacional dos métodos exatos, utilizá-los torna-se inviável para instâncias do problema de tamanho médio, justificando a utilização de heurísticas para sua resolução.

Grande parte das metaheurísticas são inspiradas em comportamentos naturais. Dentre as mais conhecidas temos a *busca tabu* [Glover and Laguna, 2003], baseada em um comportamento social, a colônia de formigas, inspirada na forma com que as formigas traçam o caminho da colônia até o alimento [Colorni et al., 1991], o algoritmo genético, este inspirado na forma com que acontece a evolução genética nos seres vivos [Holland, 1975] e o arrefecimento simulado, inspirado numa técnica de aquecimento e resfriamento controlado utilizada na produção de alguns materiais [Kirkpatrick et al., 1983].

Em 1998 foi proposto por P. Chu e J. Beasley um algoritmo genético para o MKP [Chu and Beasley, 1998]. Neste algoritmo, operadores heurísticos utilizam também conhecimento específico do problema para auxiliar no processo evolutivo. Um algoritmo de arrefecimento simulado para o MKP é proposto em [Qian and Ding, 2007]. Neste, testes computacionais mostraram que o algoritmo obteve considerável redução no tempo de execução, não alcançando porém a mesma qualidade de solução, quando comparado ao algoritmo genético proposto por P. Chu e J. Beasley.

Uma das abordagens que têm obtido melhores resultados é a abordagem híbrida proposta em [Vasquez and Hao, 2001] e aperfeiçoada em [Vasquez and Vimont, 2005]. Nessa abordagem o MKP tem o seu espaço de solução reduzido e particionado, através da adição de algumas restrições, as quais fixam o número máximo de itens a serem selecionados. Limites para estas restrições são calculados, solucionando-se uma versão relaxada do problema original. Para cada parte do espaço de busca, uma busca tabu é aplicada de forma independente, tendo como solução inicial a solução encontrada na versão relaxada do problema [Puchinger et al., 2010].

Em 2002 foi proposto por Stefka Fidanova um algoritmo de colônia de formigas para o MKP. Neste, várias informações heurísticas são utilizadas na construção das soluções [Fidanova, 2002]. Um outro algoritmo de colônia de formigas foi proposto em [Ke et al., 2008], associado-o a um procedimento de busca local. Já o trabalho apresentado em [Kong et al., 2008] propõe a resolução do MKP, utilizando uma versão do algoritmo de colônia de formigas desenvolvida especialmente para estruturas de soluções binárias. Segundo o relato do trabalho, este apresentou melhores resultados que os algoritmos de colônia de formigas convencionais aplicados ao MKP.

Em [Hanafi et al., 2010] um método híbrido é proposto para o MKP, utilizando uma abordagem de programação matemática aliada a uma busca por decomposição de vizinhança. A cada iteração, uma versão relaxada do problema é solucionada para guiar a busca por vizinhos. Segundo os testes computacionais relatados no trabalho, os resultados se comparam às heurísticas relacionadas até o momento como *estado-da-arte*.

Beliczynski Bartłomiej propôs em 2007 uma implementação da metaheurística por *exame de partículas* para o MKP, alcançando resultados próximos e, em alguns casos, iguais as melhores soluções conhecidas para as instâncias testadas [Hembeck et al., 2007].

Uma heurística híbrida, utilizando um algoritmo de estimação de distribuição em conjunto com uma busca local já foi proposta para a versão multiobjetivo do problema da mochila multidimensional [Li et al., 2004]. Segundo o trabalho, experimentos computacionais indicaram que a heurística híbrida proposta superou vários outros algoritmos, considerados até então como *estado-da-arte*. Em [Larra and Lozano, 2001] é apresentado, para fins didáticos, um algoritmo de estimação de distribuição aplicado ao problema da mochila.

Como dito anteriormente o objetivo deste projeto é propor uma heurística híbrida para a resolução do MKP, aliando a metaheurística *intelligent water drops* a um algoritmo de estimação de distribuição, esperando assim melhores

resultados que os observados em [Shah-Hosseini, 2008].

O algoritmo *intelligent water drops* é uma metaheurística de caráter populacional, ou seja, a cada passo do algoritmo, um conjunto (ou população) de soluções é gerado. É também uma heurística construtiva por construir por si só soluções completas, não necessitando de receber como entrada uma solução inicial. O IWD é inspirado na forma com que porções de águas traçam o seu caminho ao saírem de uma nascente até um ponto final onde desaguam. Se não houvessem obstáculos no caminho até o seu destino, as porções de águas em um rio fariam um caminho em linha reta, por ser o menor caminho de sua origem até o destino. Porém, devido aos vários obstáculos existentes no ambiente, o caminho real tomado pelas águas é diferente do ideal, possuindo assim voltas e torções. É interessante observar que este caminho, de aparência tortuosa, parece ser o menor quando considerados os obstáculos existentes no ambiente.

No algoritmo, o ambiente em que as gotas fluem é discreto, sendo modelado como um grafo, cujos vértices representam as soluções parciais para o problema e cada aresta, um caminho ligando duas soluções. A cada iteração do algoritmo, cada gota d'água constrói uma solução, saindo de um vértice inicial e chegando a um vértice final.

Cada aresta neste grafo possui uma quantidade de solo. Ao passar pela aresta a gota d'água retira parte do solo, carregando-a agora consigo. Cada gota d'água possui também uma velocidade: gotas d'água mais velozes retiram uma maior quantidade de solo dos caminhos em que passam. Simultaneamente, caminhos que possuem em si pouco solo acrescentam mais velocidade às gotas que por ele passam. Já caminhos que possuem muito solo acrescentam menos velocidade às gotas. Ao longo do processo as gotas tendem a escolher caminhos mais fáceis, ou seja, que possuam uma menor quantidade de solo.

Uma iteração do algoritmo termina quando todas as gotas d'água chegam numa solução completa. A melhor solução construída na iteração é então selecionada para que, a partir dela, a quantidade de solo nos caminhos sejam atualizadas. O algoritmo se encerra quando alcança uma determinada quantidade de iterações. A melhor solução encontrada é então dada como resposta.

Por sua vez, os algoritmos de estimação de distribuição (AED) são algoritmos probabilísticos que trabalham gerando populações de soluções, baseando-se em um modelo probabilístico estimado a partir de amostras de soluções selecionadas de populações anteriores [Bengoetxea, 2002]. O algoritmo abaixo descreve um

AED.

Algorithm 1: O algoritmo de estimação de distribuição.

Entrada: Número de iterações: k
Saída: Solução: s^*

```
1 início
2    $M_0 \leftarrow \text{distribuição\_uniforme}$ 
3    $D_0 \leftarrow \text{gera\_população}(M_0, R)$ 
4    $s^* \leftarrow \text{melhor\_solução}(D_0)$ 
5   para  $i \leftarrow 1$  até  $k$  faça:
6      $D_{i-1}^N \leftarrow \text{seleciona\_indivíduos}(D_{i-1}, N)$ 
7      $M_i \leftarrow \text{induz\_modelo}(D_{i-1}^N)$ 
8      $D_i \leftarrow \text{gera\_população}(M_i, R)$ 
9      $s^* \leftarrow \text{atualiza\_melhor\_solução}(D_i)$ 
10  fim
11  retorna  $s^*$ 
12 fim
```

Primeiramente o algoritmo gera uma população inicial contendo R soluções (linha 3). Esta geração inicial é geralmente feita assumindo uma distribuição uniforme sobre as variáveis que compõem a solução do problema (linha 2). Dentre esta população inicial, a melhor solução é selecionada como melhor solução global (linha 4). Num segundo passo, para gerar uma nova população, são selecionadas da população anterior N soluções (sendo $N < R$) seguindo um determinado critério (linha 6). No próximo passo, é induzido o modelo probabilístico que melhor representa as interdependências entre as variáveis da solução (linha 7). Os últimos passos são gerar uma nova população, contendo R soluções, a partir do modelo probabilístico estimado no passo anterior (linha 8), e atualizar a melhor solução global (linha 9). Geralmente, nesta nova população a melhor solução é mantida, sendo necessária a geração de apenas $R - 1$ soluções. O algoritmo então volta ao segundo passo até que o número máximo de iterações é alcançado, dando como resposta a melhor solução encontrada (linha 11).

A indução do modelo probabilístico, considerado um procedimento de aprendizagem de máquina, é o passo mais importante e crucial para o algoritmo, pois uma representação apropriada de tais interdependências é essencial para a geração de boas soluções viáveis. Assim, ao implementar um AED é importante decidir com cautela o modelo probabilístico a ser utilizado.

A vantagem do AED sobre outros métodos é registrar explicitamente as interdependências entre as diferentes variáveis que compõem a solução do problema, através de um modelo probabilístico, ao contrário da maioria das metaheurísticas, que que mantém estas interdependências de forma implícita. Este registro explícito tende a guiar o processo de busca de forma mais *inteligente*, levando a melhores soluções.

A proposta para o desenvolvimento da heurística híbrida para o MKP é utilizar um AED para auxiliar a construção das soluções geradas por cada gota d'água. Após uma primeira iteração completa do IWD, algumas melhores soluções são selecionadas da população gerada, para que um modelo probabilís-

tico seja proposto. Na próxima iteração do IWD, cada gota d'água, ao escolher os caminhos a serem tomados para a construção da solução, não só utilizaria informações do ambiente (quantidade de solo nos caminhos) mas também informações do modelo probabilístico proposto na iteração anterior. Desta forma esperamos alcançar melhores soluções para o MKP.

Outra motivação para este trabalho é aplicar a solução encontrada a um problema real da indústria de distribuição de energia elétrica brasileira. O problema em questão é a otimização de investimentos em ações de combate a perda de energia na distribuição de energia elétrica. Neste problema deve-se maximizar a recuperação de energia em um cenário plurianual de investimentos, no qual há uma disponibilidade de orçamento para cada ano do plano de investimentos. Cada ação disponível possui um custo e produz uma determinada recuperação de energia. Cada orçamento anual corresponde a uma mochila que deve ser preenchida com um conjunto de ações de modo a maximizar a energia recuperada globalmente.

4 Metodologia

O início da pesquisa se dará com uma ampla revisão bibliográfica sobre o assunto. Serão relacionados trabalhos abordando os principais métodos já propostos para a resolução do problema, bem como trabalhos abordando os métodos utilizados (metaheurística IWD e AED). Assim que alguns principais trabalhos forem relacionados, será iniciada a implementação do método híbrido proposto e dos métodos pré-existentes.

Pequenas baterias de testes seguirão a implementação com o objetivo de realizar os ajustes necessários. Finalizada a implementação e ajustes do método proposto, os resultados serão analisados para que melhorias sejam propostas.

Em seguida, serão reunidas as bases de testes utilizadas na literatura, bem como instâncias reais do problema. Os experimentos computacionais finais serão extensos, visando a melhor observação do comportamento da heurística proposta. Para isso, novas bases de testes poderão ser geradas, caso seja necessário. Os resultados obtidos no trabalho serão comparados aos resultados obtidos pelos métodos pré-existentes e aos divulgados na literatura. Uma análise estatística será feita sobre os resultados obtidos.

Assim que obtidos e analisados os resultados dos experimentos, estes serão relatados em artigos submetidos a eventos e periódicos. Os resultados parciais serão apresentados no Exame de Qualificação e na Proposta de Doutorado. A conclusão do trabalho será apresentada na Tese de Doutorado.

5 Cronograma

A Tabela 1 apresenta o cronograma do projeto de pesquisa.

Atividades	Semestres							
	1º	2º	3º	4º	5º	6º	7º	8º
Cumprimento de Créditos	●	●	●					
Revisão Bibliográfica	●	●	●					
Implementação e Ajustes	●	●	●	●	●	●		
Obtenção de Bases para Testes		●	●	●				
Testes Computacionais		●	●	●	●	●	●	
Análise dos Resultados			●	●	●	●	●	
Redação de Artigos				●	●	●	●	●
Exame de Qualificação				●				
Proposta de Doutorado						●		
Tese de Doutorado								●

Tabela 1: Cronograma do projeto de pesquisa.

Referências

- [Balas, 1965] Balas, E. (1965). An Additive Algorithm for Solving Linear Programs with Zero-One Variables. *Operations Research*, 13:517–546.
- [Bengoetxea, 2002] Bengoetxea, E. (2002). *Inexact Graph Matching Using Estimation of Distribution Algorithms*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France.
- [Blum and Roli, 2003] Blum, C. and Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, 35:268–308.
- [Chu and Beasley, 1998] Chu, P. C. and Beasley, J. E. (1998). A genetic algorithm for the multidimensional knapsack problem. *Journal of Heuristics*, 4(1):63–86.
- [Coloni et al., 1991] Coloni, A., Dorigo, M., and Maniezzo, V. (1991). Distributed Optimization by Ant Colonies. In *European Conference on Artificial Life*.
- [Dorigo, 2013] Dorigo, M. (2013). Metaheuristics network <http://www.metaheuristics.net>.
- [Fidanova, 2002] Fidanova, S. (2002). *ACO Algorithm for MKP Using Various Heuristic Information*.
- [Garey and Johnson, 1979] Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*.
- [Gavish and Pirkul, 1982] Gavish, B. and Pirkul, H. (1982). Allocation of data bases and processors in a distributed computing system.
- [Gilmore and Gomory, 1966] Gilmore, P. C. and Gomory, R. E. (1966). The Theory and Computation of Knapsack Functions. *Operations Research*, 14:1045–1074.
- [Glover and Laguna, 2003] Glover, F. and Laguna, M. (2003). Tabu Search. *Inteligencia Artificial, revista Iberoamericana De Inteligencia Artificial*, 7.
- [Hanafi et al., 2010] Hanafi, S., Lazic, J., Mladenovic, N., Wilbaut, C., and Crévits, I. (2010). *New Hybrid Metaheuristics for Solving the Multidimensional Knapsack Problem*.
- [Hembecker et al., 2007] Hembecker, F., Lopes, H., and Godoy, Walter, J. (2007). Particle swarm optimization for the multidimensional knapsack problem. In Beliczynski, B., Dzielinski, A., Iwanowski, M., and Ribeiro, B., editors, *Adaptive and Natural Computing Algorithms*, volume 4431 of *Lecture Notes in Computer Science*, pages 358–365. Springer Berlin Heidelberg.

- [Holland, 1975] Holland, J. H. (1975). Adaption in natural and artificial systems.
- [Karp, 1972] Karp, R. M. (1972). *Reducibility Among Combinatorial Problems*, pages 85–103. Plenum Press, NY.
- [Ke et al., 2008] Ke, L., Feng, Z., Ren, Z., and Wei, X. (2008). An ant colony optimization approach for the multidimensional knapsack problem. *Journal of Heuristics*, 16(1):65–83.
- [Kirkpatrick et al., 1983] Kirkpatrick, S., Vecchi, M. P., Gelatt, C. D., and kb (1983). Optimization by Simulated Annealing. In *IBM Germany Scientific Symposium Series*, volume 220, 4598, pages 671–680.
- [Kong et al., 2008] Kong, M., Tian, P., and Kao, Y. (2008). A new ant colony optimization algorithm for the multidimensional Knapsack problem. *Computers & Operations Research*, 35:2672–2683.
- [Larra and Lozano, 2001] Larra, P. and Lozano, J. A. (2001). Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation.
- [Li et al., 2004] Li, H., Zhang, Q., Tsang, E. P. K., and Ford, J. A. (2004). *Hybrid Estimation of Distribution Algorithm for Multiobjective Knapsack Problem*.
- [Lu et al., 1999] Lu, L. L., Chiu, S. Y., Cox, L. A., and Correspondence (1999). Optimal project selection: Stochastic knapsack with finite time horizon. *Journal of the Operational Research Society*, 50(6):645–650.
- [Martello and Toth, 1990] Martello, S. and Toth, P. (1990). *Knapsack Problems: Algorithms and Computer Implementations*. Wiley, New York.
- [Puchinger et al., 2010] Puchinger, J., Raidl, G. R., and Pferschy, U. (2010). The Multidimensional Knapsack Problem: Structure and Algorithms. *Informatics Journal on Computing*, 22:250–265.
- [Qian and Ding, 2007] Qian, F. and Ding, R. (2007). Simulated Annealing for the 0/1 Multidimensional Knapsack Problem.
- [Schrijver, 2003] Schrijver, A. (2003). Combinatorial optimization. Polyhedra and efficiency. *Journal of Computer and System Sciences*.
- [Shah-Hosseini, 2007] Shah-Hosseini, H. (2007). Problem solving by intelligent water drops. In *IEEE Congress on Evolutionary Computation*, pages 3226–3231.
- [Shah-Hosseini, 2008] Shah-Hosseini, H. (2008). Intelligent water drops algorithm: A new optimization method for solving the multiple knapsack problem. *International Journal of Intelligent Computing and Cybernetics*, 1:193–212.

- [Shih, 1979] Shih, W. (1979). A Branch and Bound Method for the Multiconstraint Zero-One Knapsack Problem. *Journal of The Operational Research Society*, 30:369–378.
- [Vasquez and Hao, 2001] Vasquez, M. and Hao, J. (2001). A Hybrid Approach for the 01 Multidimensional Knapsack problem. In *International Joint Conference on Artificial Intelligence*, pages 328–333.
- [Vasquez and Vimont, 2005] Vasquez, M. and Vimont, Y. (2005). Improved results on the 0-1 multidimensional knapsack problem. *European Journal of Operational Research*, 165:70–81.