

META-HEURÍSTICA VNS APLICADA AO PROBLEMA DA DIVERSIDADE MÁXIMA

Igor Lucio Rocha Alves

Universidade Federal de Viçosa

MG 230, KM 7, Caixa Postal 22, Campus UFV – Rio Paranaíba – MG

`igor.lucio@ufv.br`

RESUMO

O problema da diversidade máxima visa encontrar um subconjunto de um conjunto de forma que este seja o mais diverso possível. Este trabalho apresenta uma nova abordagem para o problema da diversidade máxima usando a meta-heurística Busca em Vizinhança Variável. Os resultados alcançados pelo algoritmo heurístico proposto, para o conjunto de instâncias MDPLIB, indicam que o algoritmo tem potencial para encontrar boas soluções, apesar de possuir um desempenho inferior aos melhores valores do conjunto de instâncias.

PALAVRAS CHAVE. Problema da Diversidade Máxima, meta-heurísticas, Busca em Vizinhança Variável

ABSTRACT

The maximum diversity problem aims to find a subset of a set so that it is as diverse as possible. This paper presents a new approach to the maximum diversity problem using the Variable Neighborhood Search metaheuristic. The results achieved by the proposed heuristic algorithm for the MDPLIB instance set indicate that the algorithm has the potential to find good solutions, although it has a performance lower than the best instance set values.

KEYWORDS. Maximum Diversity Problem, metaheuristics, Variable Neighborhood Search

1. Introdução

Este artigo descreve uma aplicação baseada no Método de Descida em Vizinhança Variável para tratar o Problema da Diversidade Máxima (PDM). O mesmo consiste em fazer a seleção de um subconjunto de m itens dentro de um conjunto de n itens, de forma que o subconjunto seja o mais diverso possível. A forma como é quantificado e qualificado a diversidade é variado, uma vez que a mesma depende do contexto do problema e sua aplicação.

Segundo Kuo et al. [1993], o PDM é classificado como NP-Difícil e limita-se ao uso exclusivo de métodos exatos para a sua solução.

Desta forma, a estrutura do artigo está organizado conforme segue.

2. Problema da Diversidade Máxima (PDM)

Segundo Moro et al. [2017], o PDM consiste em fazer a seleção de um subconjunto de M itens dentro de um conjunto de N itens, de forma que o subconjunto seja o mais diverso possível. A forma como é qualificado a diversidade é variado, uma vez que a mesma depende do contexto do problema e sua aplicação. Um exemplo para a medida de diversidade pode ser a distância de todos para todos. A distância, então, é calculada por meio dos atributos dos elementos do conjunto N .

Cada par de elementos (i, j) com $i, j \in N$ está associado a um valor d_{ij} , determinado diversidade entre os elementos i e j . Quanto maior este valor, maior a diversidade entre os elementos. Não há diversidade de um elemento para ele mesmo, isto é, $d_{ii} = 0$ e além disso, $d_{ij} = d_{ji}$. Portanto, temos uma matriz $n \times n$ espelhada (simétrica) com diagonal nula para representar a diversidade entre elementos i e j .

Diante disso, objetiva-se selecionar um subconjunto com m elementos ($m < n$), de forma que o somatório das diversidades entre eles seja o máximo possível.

Em Moro et al. [2017] há um exemplo mais detalhado da formulação do problema, bem como um estudo de uma formulação linear inteira mista para o problema da diversidade máxima.

2.1. Aplicações do Problema

O problema da diversidade máxima possui aplicações em várias áreas, como por exemplo, biologia e gerenciamento de recursos humanos.

Na área da biologia, mais precisamente em preservação ecológica, há a preocupação da conservação da diversidade biológica, devido ao aumento das taxas de degradação ambiental causadas pela intervenção humana. Embora tenha sido desenvolvido novas estratégias para a preservação de espécies em extinção, os recursos para a manutenção de programas de para tal finalidade são limitados. Diante disso, a importância na forma na aplicação de recursos de forma eficiente, uma vez que a diversidade pode ser quantificada. Glover et al. [1995] apresentaram um modelo de programação quadrática binário objetivando a maximização da biodiversidade mediante restrição de recursos.

Na área de gerenciamento de recursos humanos, duas de suas várias atribuições são a seleção de pessoal e avaliação de desempenho dos colaboradores. Supondo que determinada empresa queira realizar a seleção de três funcionários dentre cinco pré-selecionados para composição de uma equipe de trabalho e gostaria que esta tenha os mais diversos atributos possíveis. Bonotto et al. [2017] propuseram em seu trabalho um exemplo prático para este cenário, no qual utilizam uma abordagem híbrida para o problema e baseiam-se nas meta-heurísticas de otimização por nuvem de partículas e busca tabu.

3. Metodologia

Nesta etapa irei descrever a construção das heurísticas (construtiva e refinamento), função objetivo, representação da solução e por fim, a meta-heurística utilizada.

3.1. Representação da Solução

Conforme descrito na seção 2, temos uma matriz $n \times n$ simétrica com diagonal nula para a representação da diversidade dos elementos i e j . A representação será por meio de um vetor de inteiros de tamanho N onde os M primeiros serão considerados como parte da solução.

A Tabela 1 ilustra um exemplo de solução para o problema, onde os dois primeiros são os elementos escolhidos (m) e o restante dos elementos não pertence a solução; na ocasião, $n = 5$ e $m = 2$.

| | | | | |
|---|---|---|---|---|
| 0 | 3 | 1 | 4 | 2 |
|---|---|---|---|---|

Tabela 1: Representação do vetor solução

3.2. Função Objetivo

De acordo com Kuo et al. [1993], o PDM pode ser representado como um problema de programação inteira. Seja $x_i, \forall_i \in N$, se e somente se, o elemento i estiver na solução e 0, caso contrário, e d_{ij} representa a diversidade entre i e j , onde $i, j \in N$.

Diante disso, como o objetivo é maximizar a diversidade entre os elementos, temos a seguinte função objetivo:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n d(i, j) x_i x_j$$

sujeito a:

$$\sum_{i=1}^n x_i = m.$$

Este é um modelo não linear devido a sua função objetivo quadrática e tem como propósito maximizar a soma das diversidades de M elementos de N . Está sujeito a exigência de que exatos M elementos façam parte de uma solução viável.

Por questões de implementação, esta representação foi adaptada de forma que apenas os M primeiros elementos do vetor retornado seja considerado como solução, assim como ilustrado na Tabela 1.

3.3. Heurísticas Construtivas

Foi implementado 3 heurísticas construtivas: duas gulosas e uma aleatória. Onde soluções gulosas visam produzir soluções iniciais de boa qualidade, e a aleatória produzir soluções diversificadas.

3.3.1. Soma de índices

Inicialmente, cria-se um vetor que armazena todas as distâncias de todos os elementos, logo após este vetor é ordenado e os M primeiros elementos são selecionados, isto é, fazem parte da solução.

3.3.2. Semi gulosa

Sorteia-se um elemento, logo após como na soma de índices, são selecionados $M-1$ elementos.

3.3.3. Aleatória

Um vetor é preenchido com N elementos, logo após é realizado um embaralhamento.

3.4. Heurísticas de Busca Local (Refinamento)

Foi implementado 2 buscas locais afim de explorar a estrutura de vizinhança e melhorar a solução construída. São elas: Troca e Troca 2P.

3.4.1. Troca

Seleciona um elemento aleatoriamente do subconjunto M e troca com um elemento que não está no subconjunto M .

3.4.2. Troca 2P

Seleciona dois elementos aleatoriamente do subconjunto M e troca com dois elementos que não estão no subconjunto M .

As buscas locais foram incluídas no método de descida em vizinhança variável (VND) onde seu objetivo é explorar espaços de soluções através de trocas sistemáticas de estruturas de vizinhança, neste caso, troca e troca 2P.

Pelo método de primeira melhora (*First Improvement*) usa duas estruturas: Troca e Troca 2P.

Pelo método de melhor melhoria (*Best Improvement*) usa apenas a estrutura Troca.

3.4.3. Descida em Vizinhança Variável (VND)

Após a implementação dos mecanismos de busca local, tornou-se mais prático a definição do VND. Dado as vizinhanças: N_1 e N_2 .

- N_1 : Para esta vizinhança foi utilizado uma estratégia de melhor melhoria Troca 2P (*best improvement*).
- N_2 : Para explorar a vizinhança foi utilizado a estratégia de primeira melhora (Troca) (*first improvement*).

Seguimos para a definição da ordem em que tais estruturas vão ser exploradas. A vizinhança que será aplicada primeiramente é a N_1 . Sempre que houver melhorias ao aplicar outras vizinhanças, a busca será redirecionada para esta vizinhança, desta forma é garantido que o método terá fim apenas quando não for possível mais extrair melhorias de todas as vizinhanças.

3.5. Meta-heurística

De acordo com Ribeiro [1996], meta-heurísticas são procedimentos destinados a encontrar uma boa solução, eventualmente a ótima, consistindo na aplicação, em cada passo, de uma heurística subordinada, a qual tem que ser modelada para cada problema específico.

Neste caso, utilizaremos a busca em vizinhança variável (VNS) proposta por Mladenović e Hansen [1997]. VNS é uma meta-heurística de busca local que explora o espaço de soluções através de trocas sistemáticas de estruturas de vizinhança. A busca local do VNS é feita através do VND.

3.5.1. Shake

Shake é o movimento utilizado para perturbação da solução onde seleciona metade da solução, realiza a troca (ou troca 2P) e sorteia elementos que não estão na solução para serem incluídos.

4. Resultados

As instâncias utilizadas para a validação da abordagem descrita neste trabalho é a MDPLIB, de autoria do Optsicom [2016]. A MDPLIB contém um compilado de 315 instâncias para o problema da diversidade máxima.

A seguir, é apresentada uma breve descrição da origem e das características do conjunto de instâncias:

- **SOM**: consiste em 70 matrizes com números aleatórios entre 0 e 9 gerados a partir de uma distribuição uniforme inteira.
- **GKD**: consiste em 145 matrizes para as quais os valores foram calculados como as distâncias euclidianas dos pontos gerados aleatoriamente com coordenadas no intervalo de 0 a 10.
- **MDG**: consiste em 100 matrizes com números reais selecionados aleatoriamente entre 0 e 10 de uma distribuição uniforme.

4.1. Parâmetros

É necessário passar, por parâmetro, a taxa do *shake* e o número de iterações. A taxa do *shake* será responsável por mensurar a perturbação da solução, e o número de iterações como critério de parada.

4.2. Testes

Os parâmetros utilizados foram os seguintes:

- Taxa *shake*: 50%
- Número de Iterações: 20.

Com base nos parâmetros definidos acima, foram realizados 4 testes utilizando 4 instâncias dos 3 conjuntos (SOM, MDG e GKD); e comparados com os melhores valores do Optsicom [2016]. A Tabela 2 mostra os resultados, sendo a coluna *VNS* o valor alcançado pela meta-heurística utilizada neste trabalho.

| Instância | MDPLIB | VNS |
|-------------------|-----------|---------|
| MDG-a_1_n500_m50 | 7833,83 | 7828,4 |
| SOM-b_5_n200_m20 | 1247 | 1243 |
| MDG-b_1_n500_m50 | 778030,62 | 774598 |
| GKD-c_20_n500_m50 | 19604,84 | 19604,8 |

Tabela 2: MDPLIB \times VNS

Os algoritmos neste trabalho foram implementados sobre a linguagem de programação C++ e avaliados em um *hardware* composto por: processador Intel Core™ i7-8550U, 16GB de RAM e 240GB de SSD (*Solid State Drive*).

5. Conclusão

A abordagem proposta neste trabalho para encontrar soluções de boa qualidade para o problema em questão permitiu a exploração de um extenso espaço de busca, evidenciando resultados satisfatórios. A utilização de taxas de perturbação na exploração levou a proporcionar uma maximização do espaço de busca.

Como observado na tabela 2 da seção anterior, os resultados são, em maioria, satisfatórios, mesmo que para algumas instâncias não tenha igualado ou ultrapassado do Optsicom [2016].

Um trabalho futuro será implementar novas estratégias com base em outras meta-heurísticas. Outro ponto é na forma de avaliar o tempo de execução do algoritmo, objetivando atenuá-lo, mesmo diante de uma alta carga de trabalho.

Referências

- Bonotto, E. L. et al. (2017). Otimização por nuvem de partículas e busca tabu para problema da diversidade máxima.
- Glover, F., Ching-Chung, K., e Dhir, K. S. (1995). A discrete optimization model for preserving biological diversity. *Applied mathematical modelling*, 19(11):696–701.
- Kuo, C.-C., Glover, F., e Dhir, K. S. (1993). Analyzing and modeling the maximum diversity problem by zero-one programming. *Decision Sciences*, 24(6):1171–1185.
- Mladenović, N. e Hansen, P. (1997). Variable neighborhood search. *Computers & operations research*, 24(11):1097–1100.
- Moro, M. A. et al. (2017). Meta-heurísticas grasp e brkga aplicadas ao problema da diversidade máxima.
- Opticom, P. (2016). Mdplib. <http://grafo.etsii.urjc.es/opticom/mdp>. Acesso: 06/12/2019.
- Ribeiro, C. C. (1996). Metaheuristics and applications. *Advanced School on Artificial Intelligence, Estoril, Portugal*.