UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE EDUCAÇÃO SUPERIOR DO ALTO VALE DO ITAJAÍ –
CEAVI DEPARTAMENTO DE ENGENHARIA DE SOFTWARE - DESO
Breno josé Cooelho (breno.jose.coelho@gmail.com)

ALGORITMOS HEURÍSTICOS PARA O PROBLEMA DA DIVERSIDADE MÁXIMA

1. O PROBLEMA

Neste estudo, o foco se direciona para o uso do Problema de Diversidade Máxima (PDM), o PDM é um problema de otimização que consiste em selecionar um subconjunto de elementos de um conjunto maior de modo que a diversidade entre os elementos selecionados seja maximizada. A diversidade é medida pela distância entre os elementos. O PDM tem muitas aplicações, como seleção de portfólio, seleção de produtos e seleção de funcionários.

Foi desenvolvido o modelo de Problema de Programação Linear Inteira no qual foi descrita . Função Objetivo:

Maximizar

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

sujeito a:

$$\sum_{i=1}^{n} x_i = m, \tag{5}$$

$$x_i + x_j - y_{ij} \leq 1, \forall (i,j) \in Q, \tag{6}$$

$$-x_i + y_{ij} \leq 0, \forall (i,j) \in Q, \tag{7}$$

$$-x_j + y_{ij} \leq 0, \forall (i,j) \in Q, \tag{8}$$

$$y_{ij} \in \{0,1\}, \forall (i,j) \in Q, \tag{9}$$

$$x_i \in \{0,1\}, \forall i \in N. \tag{10}$$

Em resumo, o modelo matemático descrito busca encontrar um subconjunto de elementos que maximize a diversidade total. As variáveis x_i determinam quais elementos fazem parte da solução, enquanto as variáveis y_ij indicam quais pares de elementos contribuem para a diversidade total.

Restrições:

Essa restrição (5) garante que exatamente m elementos sejam escolhidos para a solução. Em outras palavras, a soma de todas as variáveis x_i deve ser igual a m.

Temos a restrição (6) que define a relação entre as variáveis x_i e y_ij. Elas garantem que y_ij só será 1 se tanto x_i quanto x_j forem 1. Ou seja, y_ij só será 1 se os elementos i e j forem ambos selecionados.

As restrições (7) e (8) complementam as restrições (6). Elas garantem que se um elemento i (ou j) foi selecionado (x_i = 1), então existe pelo menos um outro elemento j (ou i) tal que o par (i,j) foi selecionado (y_ij = 1). Isso significa que cada elemento selecionado deve contribuir para a diversidade total da solução.

Ja as restrições (9) e (10) simplesmente reafirmam a natureza binária das variáveis x_i e y_ij, respectivamente.

- xi = Variável binária que assume o valor 1 se o item ii for incluído na solução, e 0 caso contrário. Esse valor é restrito a {0,1} para i ∈ N.

- yij = Variável binária que assume o valor 1 se uma certa relação entre os itens i e j for satisfeita e 0 caso contrário. Assim como xi, yij ∈ {0,1} para todo par (i,j) ∈ Q.

- dij: Parâmetro constante que representa o peso ou valor associado ao par de itens (i,j) quando ambos estão incluídos, utilizado na função objetivo.

Detalhes sobre o problema e sua modelagem, estão disponíveis no artigo usado como base disponibilizado no repositório.

2. INSTÂNCIAS

Os artigo utilizado na pesquisa ofereceu instâncias que poderiam ser utilizadas, porém o desempenho foi ruim para casos de demonstração estudo, levou horas e não gerou algum resultado, então as instancias possuem o valo de (N) representando o número total de elementos no conjunto, sendo a dimensão da matriz de custos d_ij. Cada linha e coluna dessa matriz corresponde a um elemento do conjunto. E o valor de (M) representando o número de elementos que devem ser selecionados do conjunto total, o objetivo é escolher M elementos de forma que a soma das dissimilaridades entre eles seja máxima.

Exemplo do corpo de uma instância:

10 2

0 1 166.47234

0 2 170.18475

0 3 174.55453

0 4 153.28670

0 5 145.12186

0 6 144.42723

0 7 170.98466

0 8 176.58110

0 9 162.99632

1 2 168.48360

1 3 175.63185

1 4 135.43608

1 5 153.11269

1 6 147.68130

1 7 195.46092

1 8 156.55237

1 9 184.35828

2 3 183.43506

2 4 159.96216

2 5 150.25894

2 6 131.15147

2 7 169.92273

2 8 168.09035

2 9 193.29750

3 4 187.82150

3 5 122.49923

3 6 196.62941

3 7 182.58435

3 8 144.85728

3 9 187.11629

4 5 167.54440

4 6 162.23880

4 7 176.99178

4 8 196.17409

4 9 127.35979

5 6 140.80400

5 7 108.77892

5 8 131.51010

5 9 165.05664

6 7 162.18561

6 8 199.05042

6 9 173.73540

7 8 164.31947

7 9 188.75363

8 9 243.97252

3. IMPLEMENTAÇÃO

Disponível em:

https://github.com/BrenoJoseCoelho/ProblemaDiversidadeMaximaMetodosQuantitati vos

3.1 LEITURA DE INSTÂNCIA

No código, a leitura das instâncias é feita pela função read_instance(file_path), que começa abrindo o arquivo de entrada e lendo todas as suas linhas. A primeira linha contém os valores de N (número de nós) e M (número de nós a serem selecionados), que são extraídos e convertidos para inteiros. A partir da segunda linha, as distâncias entre os pares de nós são lidas. Cada linha contém três valores: os índices i e j dos nós e a distância entre eles (d(i,j)). Esses valores são convertidos para inteiros e a distância para um valor de ponto flutuante. A matriz de distâncias d_ij é preenchida de forma simétrica, ou seja, d(i,j) é igual a d(j,i), e essa matriz é retornada juntamente com os valores de N e M. Esses dados são então utilizados no modelo de otimização para definir a função objetivo e as restrições.

5. RESULTADOS

Após a execução do projeto, no qual os valores das 10 instâncias foram extraídas no artigo, aqui estão os resultado de valor ótimo e o tempo de execução:

Instância	Valor Ótimo	Tempo
1	243.97	0.23 /s
2	210.65	0.07 /s
3	363.24	0.07 /s
4	541.08	0.10 /s
5	649.72	0.25 /s
6	1181.47	0.07 /s
7	2106.73	0.07 /s

8	2349.91	0.08 /s
9	2720.61	0.11 /s
10	3506.08	0.06 /s

6. CONCLUSÃO

O presente estudo focou no desenvolvimento e aplicação de um modelo de Programação Linear Inteira (PLI) para o Problema da Diversidade Máxima (PDM), visando a seleção de subconjuntos de elementos de maneira a maximizar a diversidade entre eles. O modelo matemático foi capaz de representar com precisão as restrições do problema e a função objetivo, proporcionando uma solução viável para as instâncias fornecidas.

Os resultados obtidos nas instâncias testadas indicaram que o modelo conseguiu encontrar valores ótimos consistentes, com um bom desempenho em termos de tempo de execução. O tempo médio para a resolução das instâncias foi de poucos segundos, variando de 0.06 a 0.25 segundos, o que demonstra a eficiência do método para as instâncias de tamanho moderado. O valor da função objetivo também apresentou uma variação significativa, refletindo as diferenças nas instâncias em termos de complexidade e densidade das distâncias entre os elementos. As instâncias com maior valor de diversidade, como a instância 10 (com valor ótimo de 3506.08), mostraram-se mais desafiadoras, mas ainda assim foram resolvidas em tempos muito rápidos.

Esses resultados são promissores, pois indicam que o modelo baseado em PLI é capaz de lidar com a complexidade do PDM, proporcionando soluções ótimas em tempo hábil. Esse desempenho pode ser crucial em cenários do mundo real onde a rapidez na tomada de decisão é essencial, como na seleção de portfólios financeiros ou em processos de alocação de recursos.

Embora o modelo tenha se mostrado eficiente para as instâncias testadas, ele ainda pode ser aprimorado para lidar com instâncias de maior escala, possivelmente por meio da implementação de algoritmos heurísticos que possam fornecer soluções aproximadas em menos tempo. Em suma, os resultados obtidos não apenas validaram a eficácia do modelo para o PDM, mas também abriram caminho para novas abordagens e otimizações, especialmente para problemas e maior complexidade e tamanho.

REFERÊNCIAS

Repositório GitHub:https://github.com/BrenoJoseCoelho/ProblemaDiversidadeMaximaMetodosQuantitativos

Duarte, I. L., Silva, G. C., & Costa, T. A. (2008). **Algoritmos heurísticos para o problema da diversidade máxima**. Pesquisa Operacional, 28(3), 499-516.