

Concepção e implementação de algoritmos baseados em espaços de formulações para o agrupamento de dados

Lucas Lyon Azevedo
Éverton Santi

Congresso de Iniciação Científica e Tecnológica (CICT)
Universidade Federal do Rio Grande do Norte

OBJETIVO

O principal objetivo deste trabalho é resolver o problema de *clustering* heterogêneo utilizando um Algoritmo de Busca em Espaços de Formulações (FSS)

Problema de *Clustering* Heterogêneo (PCH)

- ▶ Baseado no conceito de medianas para particionar os objetos;
- ▶ Considera múltiplas matrizes de dissimilaridades como entrada;
- ▶ Redução da perda de informação dado a eliminação do processo de agregação;
- ▶ Santi E., Aloise D., and Blanchard S. J. A model for clustering data from heterogeneous dissimilarities. 2016.

Algoritmo Proposto (FSS) - Métodos

Algoritmos de Busca em Espaços de Formulações (FSS)

- ▶ Um problema de otimização pode ter mais de uma formulação;
- ▶ Uma solução pode ser ponto estacionário em uma formulação, mas não necessariamente o é em outra;
- ▶ Alternar entre formulações para escapar de ótimos locais;
- ▶ Diferentes formulações podem levar a diferentes soluções;

- ▶ Testou-se o algoritmo proposto por López (2014) para diversas formulações PCH;
- ▶ Resultados mostraram inviabilidade de aplicação ao PCH;
- ▶ Problema NP-difícil;
- ▶ Instâncias de grande porte;
- ▶ Elevado número de variáveis de decisão.

Algoritmo Proposto (FSS)

Dado o problema original P , obter um problema auxiliar P^* :

- ▶ Remover restrições de integralidade sobre as variáveis de decisão;
- ▶ Adicionar perturbações à formulação por meio de uma única restrição;

Ideia geral:

- ▶ Alternar entre resolver o problema P^* e recuperar uma solução viável para P .

Algoritmo Proposto

$z_{best} \leftarrow \infty \quad \delta \leftarrow 0.5 \quad \beta \leftarrow 0.9$

while *Condição de parada não atingida* **do**

Resolva P^* utilizando algum resolvidor não linear

Arredonde os valores das variáveis e para o inteiro mais próximo

do

Resolva o problema P considerando os valores das variáveis e como fixos, atribuindo o custo da solução a z_1

Resolva o problema P considerando os valores das variáveis z como fixos, atribuindo o custo da solução a z_2

while $z_1 \neq z_2$

Atualize z_{best} caso uma solução viável e de melhor custo tenha sido encontrada

Atualize δ como $\delta \leftarrow \beta\delta$

end

Resultados

- ▶ 27 Instâncias da literatura
- ▶ Simon J. Blanchard, Daniel Aloise, and Wayne S. Desarbo. The hetero-geneous p-median problem for categorization based clustering. 2012
- ▶ Custo da solução encontrada pelo FSS é pior em 19 casos, igual em 7 e melhor em 1;
- ▶ 3 soluções ótimas;
- ▶ Diferença de custos entre soluções do VNS e FSS é, em média, de 3%;
- ▶ Tempo de execução do FSS corresponde, em média, a 6% do tempo demandado pelo VNS;
- ▶ Nenad Mladenovic, Frank Plastria, and Dragan Urošević. Reformulation descent applied to circle packing problems. 2005

Considerações finais

- ▶ FSS é competitivo com VNS;
- ▶ É viável de aplicação, considerando os propósitos do PCH;
- ▶ Custo computacional relativamente baixo;
- ▶ Refinamento do algoritmo proposto;

Bibliografia



É. Santi , D. Aloise e S. J. Blanchard

A model for clustering data from heterogeneous dissimilarities.
European Journal of Operational Research, 253:659–672
(2016).



C. O. López , J. E. Beasley.

A formulation space search heuristic for packing unequal circles
in a fixed size circular container.
European Journal of Operational Research, 251:64–73, (2016).



C. O. López e J. E. Beasley

A note on solving MINLP's using formulation space search.
Optimization Letters, 8:1167–1182, (2014).



Nenad Mladenovic, Frank Plastria e Dragan Urošević

A note on solving MINLP's using formulation space search.
Optimization Letters, 8:1167–1182, (2014).