

Um modelo por programação inteira mista com meta-heurísticas para o problema de agrupamento por correlação com sobreposição de clusters

Guilherme Oliveira Chagas; Luiz Antonio Nogueira Lorena; Rafael Duarte Coelho dos Santos

guilherme.chagas@inpe.br

Problemas de agrupamento são oriundos de várias áreas da ciência. Em geral, considerando um conjunto de dados, deve-se gerar subconjuntos disjuntos desses dados de modo que dados de um mesmo subconjunto sejam similares entre si e dados de subconjuntos diferentes sejam diferentes entre si. Porém, há casos em que é necessário que esses subconjuntos possuam sobreposições, ou seja, não sejam disjuntos. Dessa maneira, deseja-se que dados semelhantes pertençam ao mesmo conjunto de subconjuntos e dados que sejam diferentes pertençam a quase nenhum, ou nenhum, subconjunto em comum. No contexto de teoria dos grafos, dado um grafo $G=(V, E)$ com arestas valoradas com "+" (positivas) e "-" (negativas), o objetivo é particionar os vértices de V em subconjuntos em que o número de arestas positivas entre vértices pertencentes a um mesmo subconjunto seja maximizado e o número de arestas negativas entre vértices que pertençam a subconjuntos diferentes seja minimizado. Realizar esse particionamento dos vértices de um grafo é conhecido como o problema de agrupamento por correlação com sobreposição. Neste trabalho é proposto um método para gerar agrupamentos por correlação com sobreposição dos vértices de um grafo. Utiliza-se, nesse método, duas meta-heurísticas para gerar várias soluções de agrupamentos por correção sem sobreposição, ou seja, subconjuntos disjuntos de vértices. Em seguida, os agrupamentos de cada solução gerada são armazenados. Com isso, um modelo por programação inteira mista é resolvido, pelo CPLEX, utilizando esses agrupamentos sem sobreposição para encontrar um agrupamento por correlação com sobreposição.

Agrupamento por correlação com sobreposição. edição de clusters com sobreposição.