

Metaheurística GRASP Aplicada ao Problema de Clusterização Hierárquica pelo Método Single Linkage

Napoleão Póvoa Ribeiro Filho

Mestrado em Modelagem Computacional de Sistemas - Universidade Federal do Tocantins
Av. NS 15, S/N – ALCNO 14 - Bloco Bala 2 – Sala 01, CEP: 77020-210, Palmas - TO
napoleao@ifto.edu.br

Marcelo Lisboa Rocha

Mestrado em Modelagem Computacional de Sistemas - Universidade Federal do Tocantins
Av. NS 15, S/N – ALCNO 14 - Bloco Bala 2 – Sala 01, CEP: 77020-210, Palmas - TO
marcelolisboarochoa@gmail.com

David Nadler Prata

Mestrado em Modelagem Computacional de Sistemas - Universidade Federal do Tocantins
Av. NS 15, S/N – ALCNO 14 - Bloco Bala 2 – Sala 01, CEP: 77020-210, Palmas - TO
ddnprata@gmail.com

RESUMO

Este trabalho traz uma nova proposta para melhora da clusterização hierárquica pelo método *single linkage*, utilizando como parâmetro o coeficiente de correlação copenética (ccc). Na clusterização hierárquica, parte-se de uma matriz de dissimilaridades e utilizando como critério a menor distância entre os pontos, onde os mesmos vão sendo agrupados até gerar ao final uma árvore invertida conhecida como dendrograma. O ccc, obtido após a construção do dendrograma, indica o quão fiel está o agrupamento em relação a matriz de dados originais. Com o objetivo de gerar dendrogramas que resultem em melhores ccc, propõe-se no presente trabalho um novo algoritmo que utiliza os conceitos da metaheurística GRASP. Testes foram realizados para comprovar o desempenho do algoritmo proposto, comparando os coeficientes obtidos com os gerados pelo software R.

Palavras Chave: GRASP, Clusterização Hierárquica, Coeficiente de Correlação Copenética

ABSTRACT

This paper presents a new proposal for improved hierarchical clustering by single linkage method, using as parameter copenetic correlation coefficient (ccc). In hierarchical clustering, it's used a dissimilarity matrix and using as a criterion the shortest distance between points, which are grouped and at the end generate an inverted tree known as dendrogram. The ccc, after the construction of the dendrogram indicates how faithful is the grouping as regards the original data matrix. In order to generate dendrograms that result in better ccc, is propose in this paper a new algorithm that uses the concepts of GRASP. Tests were conducted to confirm the performance of the proposed algorithm by comparing the coefficients obtained with those generated by software R.

Key Words: GRASP, Hierarchical clustering, Copenetic Correlation Coefficient