

Resumo

A Análise Classificatória é uma área da Análise de Dados Multivariados cujo objectivo é agrupar um conjunto de objectos a classificar num número pequeno de classes, que reflitam as relações de semelhança e/ou oposição entre esses objectos.

Um método de Classificação produz, sobre estes objectos, um conjunto de classes organizado segundo uma determinada estrutura. Duas metodologias de Classificação serão analisadas: a Classificação Hierárquica e a Classificação Piramidal.

A aplicação de um método de Classificação Piramidal produz estruturas mais complexas do que a aplicação de um método de Classificação Hierárquica. Assim, inicialmente será introduzido um vasto conjunto de definições e propriedades, apresentados os principais algoritmos de Classificação Piramidal Ascendente, serão referidas as suas características e limitações.

A motivação para desenvolver algoritmos de geração de estruturas classificatórias tem-se intensificado entre a comunidade científica, tendo dado origem a diversos estudos. Neste trabalho propõe-se um método de geração aleatória de estruturas piramidais. Sendo o modelo de Classificação Piramidal uma generalização do modelo de Classificação Hierárquica, o método aqui proposto poderá ser visto como uma extensão de trabalhos anteriores de geração aleatória de dendrogramas.

Tendo por objectivo avaliar o desempenho do método proposto, em particular, verificar se o método gera pirâmides aleatória e uniformemente, foi previamente realizado um estudo topológico de pirâmides. Para um número fixo de nós terminais identificam-se os diferentes tipos topológicos de pirâmides, assim como o respectivo número de pirâmides não isomórficas. Em virtude da elevada complexidade desta análise, este objectivo foi apenas conseguido para um número reduzido de nós terminais.

Abstract

The Cluster Analysis is an area of the Multivariate Data Analysis whose the main goal is to group a set of elements in a small number of classes, that they reflect the

relationships of similarity and/or opposition among those elements. A method of Classification produces, on these elements, a group of classes organized second a certain structure. Two methodologies of Classification will be analyzed: the Hierarchical Classification and the Pyramidal Classification.

The application of a Pyramidal Classification method produces more complex structures than the application of a Hierarchical Classification method. First, it will be introduced a vast group of definitions and properties and presented the main sort algorithms of Ascendant Pyramidal Classification.

The motivation to develop generation algorithms of classification structures have been intensifying among the scientific community, having created several studies.

In this work a random generation method of pyramidal structures is proposed.

Being the model of Pyramidal Classification a generalization of the model of Hierarchical Classification, the method proposed appears as an extension of previous works in random generation of dendrograms.

With the purpose to evaluate the performance of the proposed method, in particular, to verify if the method generates random and uniform pyramids, it was previously accomplished a topologic study of the pyramids. For a fixed number of terminal nodes the different topologic types of pyramids were identified, as well as the respective number of non isomorphic pyramids. On account of the high complexity of this analysis, this purpose was just gotten for a reduced number of terminal nodes.