PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Marcos Neves do Vale

Agrupamentos de Dados: Avaliação de Métodos e Desenvolvimento de Aplicativo para Análise de Grupos

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio.

Orientadores: Marley M. B. R. Vellasco
Ricardo Tanscheit

Rio de Janeiro, agosto de 2005

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Marcos Neves do Vale

Agrupamentos de Dados: Avaliação de Métodos e Desenvolvimento de Aplicativo para Análise de Grupos

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Marley M. B. R. Vellasco
Orientadora
PUC-RIO

Ricardo Tanscheit
Orientador
PUC-RIO

Juan Guillermo Lazo Lazo
PUC-RIO

Flávio Joaquim de Souza
UERJ

José Eugenio Leal
Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 29 de agosto de 2005

Marcos Neves do Vale

Graduou-se em Engenharia Elétrica com Ênfase em Sistemas e Computação na UERJ em 2001 e concluiu o curso de pós-graduação em Análise, Projeto e Gerência de Sistemas na PUC-RJ em 2003.

Ficha Catalográfica

Vale, Marcos Neves do

Agrupamentos de dados : avaliação de métodos e desenvolvimento de aplicativo para análise de grupos / Marcos Neves do Vale ; orientadores: Marley M. B. R. Vellasco, Ricardo Tanscheit. – Rio de Janeiro : PUC, Departamento de Engenharia Elétrica, 2005.

120 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica

Inclui referências bibliográficas.

1. Engenharia elétrica – Teses. 2. Agrupamento. 3. Análise de agrupamentos. 4. Aplicativo. 5. Lógica fuzzy. 6. Extração de informação. I. Vellasco, Marley M. B. R. II. Tanscheit, Ricardo. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Aos meus pais, principalmente a minha esposa e a minha filha.

Agradecimentos

Aos meus orientadores professora Marley Vellasco e professor Ricardo Tanscheit pela confiança, estímulo e dedicação para a realização deste trabalho.

Aos meus amigos Douglas e Paulo Motta por todo apoio, paciência e pelas importantes contribuições.

Ao Rodrigo Simões por ter me ajudado, mesmo de longe.

A todos os amigos e familiares que de uma forma ou de outra me estimularam ou me ajudaram.

Resumo

Vale, Marcos Neves do. **Agrupamentos de Dados: Avaliação de Métodos e Desenvolvimento de Aplicativo para Análise de Grupos.** Rio de Janeiro, 2005. 120p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A enorme massa de dados que é gerada pelas diversas empresas diariamente pode conter informações importantes que não são fáceis de serem extraídas. Com isso advém a necessidade de analisá-los automaticamente, de forma adequada, extraindo informação útil que pode agregar algum tipo de conhecimento. Uma das formas de se analisar os dados automaticamente é através da análise de agrupamentos. Ela procura encontrar grupos de dados semelhantes entre si. As técnicas de análise de agrupamentos revelam como os dados estão estruturados e resultam em um melhor entendimento sobre o negócio. Existe ainda hoje uma escassez de ferramentas para esse fim. Em um problema real de agrupamento de dados convém analisar os dados através da utilização de diferentes métodos, a fim de buscar aquele que melhor se adapte ao problema. Porém, as ferramentas existentes hoje em dia não são integradas, onde cada ferramenta possui um subconjunto dos métodos existentes de agrupamento. Dessa forma o usuário fica limitado à utilização de uma ferramenta específica ou é obrigado a conhecer diversas ferramentas diferentes, de forma a melhor analisar os dados de sua empresa. Esta dissertação apresenta uma revisão detalhada de todo o processo de análise de agrupamentos e o desenvolvimento de um aplicativo que visa não apenas a atender as deficiências presentes na maioria das ferramentas com esse fim, mas também a auxiliar, de forma mais completa, todo o processo de análise dos grupos. O aplicativo desenvolvido é de fácil utilização e permite que a ele sejam incorporados outros métodos eventualmente desenvolvidos pelo usuário. O aplicativo foi avaliado em três estudos de casos, os quais visam demonstrar a facilidade de uso do aplicativo, assim como avaliar as vantagens do uso de métodos de natureza fuzzy em uma base de dados real.

Palavras-chave

Agrupamento, análise de agrupamentos, aplicativo, lógica fuzzy, extração de informação.

Abstract

Vale, Marcos Neves do. **Data Clustering: Analysis of Methods and Development of Application for Cluster Analysis.** Rio de Janeiro, 2005. 120p. MSc. Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The enormous data mass that is daily generated by several companies can contain critical information that might not be easily retrieved, considering that the amount of data is generally huge and/or the target information might be spread through different data bases. Taking that into consideration, it might be necessary to properly analyze the data in an automatic way, so useful and valuable information can be extracted. One way of automatically analyzing data is through cluster analysis. This type of analysis searches for related similar data. These clusters settle a data structure model and with proper analysis can reveal important information. The techniques used in cluster analysis disclose how data is structured and allow a better knowledge of the business. Still today there is a lack of tools for this purpose. On a real situation with a data cluster problem it is wise to analyze the data through different methods, so we can find the one that better fits the problem. However, today the existing tools are not integrated, and each tool has a subgroup of existing cluster methods. This way the user stays limited to use only one specific tool or is forced to be aware of a number of different tools, so he would be able to better analyze the company data. This study presents a detailed review of the whole group analysis process and develops an application that not only suggests how to cover the currently lack of tools for this purpose, but also to help the complete cluster analysis process in a more extended way. The application developed is user friendly and allows other methods developed by users to be incorporated. The application has been evaluated into three case studies with the purpose of demonstrating its user friendly, as well as evaluating the advantages of using fuzzy methods on a true data base.

Keywords

clustering, clustering software, fuzzy clustering, information retrieval

Sumário

Lista de figuras

Lista de tabelas