Pontifícia Universidade Católica
do Rio de Janeiro

Rodrigo Tosta Peres

Novas Técnicas de Classificação de Padrões baseadas em Métodos Local-Global

Tese de Doutorado

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

Orientador: Carlos Kubrusly

Rio de Janeiro, agosto de 2008

Pontifícia Universidade Católica
do Rio de Janeiro

Rodrigo Tosta Peres

Novas Técnicas de Classificação de Padrões baseadas em Métodos Local-Global

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

Dr. Carlos Kubrusly
Orientador
Departamento de Engenharia Elétrica/PUC-Rio

Dr. Raul Queiroz Feitosa
Departamento de Engenharia Elétrica/PUC-Rio

Dr. Marcelo da Cunha Medeiros
Departamento de Economia/PUC-Rio

Dr. Amit Bhaya
UFRJ

Dr. Marcello Luiz Rodrigues Campos
COPPE/UFRJ

Dr. Sérgio Lima Netto
COPPE/UFRJ

Dr. Alexandre P. Alves da Silva
UFRJ

Prof. José Eugênio Leal
Coordenador Setorial do Centro
Técnico Científico - PUC-Rio

Rio de Janeiro 11 de agosto de 2008

Rodrigo Tosta Peres

Possui graduação em Licenciatura em Matemática pela Universidade Federal Fluminense (2000) e mestrado em Engenharia Elétrica pela Pontifícia Universidade Católica do Rio de Janeiro (2004). Tem experiência na área de Inteligência Computacional, com aplicações em Classificação de Padrões.

Ficha Catalográfica

Peres, Rodrigo Tosta

Novas técnicas de classificação de padrões baseadas em métodos local-global / Rodrigo Tosta Peres ; orientador: Carlos Kubrusly. – 2008.

81 f. : Il. ; 30 cm

Tese (Doutorado em Engenharia Elétrica)– Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Classificação de padrões. 3. Quantização vetorial. 4. Teoria da informação. 5. Máquinas de vetor de suporte. I. Kubrusly, Carlos. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Agradecimentos

A Deus, que sempre nos garante a certeza da vitória.

Ao meu orientador, Professor Carlos Eduardo Pedreira, por tudo que vem me ensinando ao longo desses anos de trabalho e pela parceria e amizade que permanecerão para sempre.

Ao professor Carlos Kubrusly, pela amizade e confiança de que este trabalho seria bem sucedido.

A PUC-Rio e ao CNPq, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos meus pais, Mário e Suely, e aos meus irmãos, Diego e Tiago, por terem me dado à base familiar que é tão importante na vida.

Ao professor Alexandre Pinto Alves da Silva, por todas as contribuições que foram muito importantes para o desenvolvimento desta tese.

Ao professor Moisés Henrique Szwarcman, pela amizade e por toda a ajuda administrativa na reta final do trabalho.

Aos amigos Hélio Pinto, André Sih e Francisco Carlos de Azevedo Pinto pelas contribuições dadas.

A banca desta tese por todo o trabalho de revisão e todas as sugestões que em muito contribuíram para este trabalho.

A todos os familiares, amigos e colegas que, direta ou indiretamente, contribuíram para que este trabalho pudesse ser desenvolvido.

Resumo

Peres, Rodrigo Tosta; Kubrusly, Carlos. **Novas Técnicas de Classificação de Padrões baseadas em Métodos Local-Global.** Rio de Janeiro, 2008. 81p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O foco desta tese está direcionado a problemas de Classificação de Padrões. A proposta central é desenvolver e testar alguns novos algoritmos para ambientes supervisionados, utilizando um enfoque local-global. As principais contribuições são: (i) Desenvolvimento de método baseado em quantização vetorial com posterior classificação supervisionada local. O objetivo é resolver o problema de classificação estimando as probabilidades posteriores em regiões próximas à fronteira de decisão; (ii) Proposta do que denominamos 'Zona de Risco Generalizada', um método independente de modelo, para encontrar as observações vizinhas à fronteira de decisão; (iii) Proposta de método que denominamos 'Quantizador Vetorial das Fronteiras de Decisão', um método de classificação que utiliza protótipos, cujo objetivo é construir uma aproximação quantizada das regiões vizinhas à fronteira de decisão. Todos os métodos propostos foram testados em bancos de dados, alguns sintéticos e outros publicamente disponíveis.

Palavras-chave

Classificação de Padrões; Quantização Vetorial; Seleção de Observações; Teoria da Informação; Máquinas de Vetor de Suporte.

Abstract

Peres, Rodrigo Tosta; Kubrusly, Carlos. **New Techniques of Pattern Classification based on Local-Global Methods.** Rio de Janeiro, 2008. 81p. PhD.Thesis - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis is focused on Pattern Classification problems. The objective is to develop and test new supervised algorithms with a local-global approach. The main contributions are: (i) A method based on vector quantization with posterior supervised local classification. The classification problem is solved by the estimation of the posterior probabilities near the decision boundary; (ii) Propose of what we call 'Zona de Risco Generalizada', an independent model method to find observations near the decision boundary; (iii) Propose of what we call 'Quantizador Vetorial das Fronteiras de Decisão', a classification method based on prototypes that build a quantized approximation of the decision boundary. All methods were tested in synthetics or real datasets.

Keywords

Pattern Classification; Vector Quantization; Observations Selection; Information Theory; Support Vector Machines.

Sumário

Lista de figuras

Lista de tabelas

Ora, àquele que é poderoso para fazer tudo muito mais abundantemente além daquilo que pedimos ou pensamos, segundo o poder que em nós opera, a esse glória na igreja, por Jesus Cristo, em todas as gerações, para todo o sempre. Amém!

(Efésios 3:20-21)