PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Thiago Baptista Rodrigues

Seleção de Variáveis e Classificação de Padrões por Redes Neurais como Auxílio ao Diagnóstico de Doença Cardíaca

Dissertação de Mestrado

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

Orientador: Prof. Carlos Kubrusly

Co-Orientador: Prof. José Leonardo Ribeiro Macrini

Rio de Janeiro

Dezembro de 2006

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Thiago Baptista Rodrigues

Seleção de Variáveis e Classificação de Padrões por Redes Neurais como Auxílio ao Diagnóstico de Doença Cardíaca

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

Prof. Carlos S. Kubrusly
Orientador
Departamento de Engenharia Elétrica – PUC-Rio

Prof. José Leonardo Ribeiro Macrini
Co-Orientador
Departamento de Metrologia

Prof. Elisabeth Costa Monteiro
Departamento de Metrologia

Prof. David Sérgio Adães de Gouvêa
UFJF

Prof. Maurício Nogueira Frota
Departamento de Metrologia

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 01 de dezembro de 2006

Thiago Baptista Rodrigues

Nascido em Vassouras-RJ em 1982. Graduou-se em Engenharia Elétrica (2004) pela Universidade Federal de Juiz de Fora, UFJF. Suas pesquisas de interesse incluem as áreas de sistemas inteligentes aplicados à classificação de padrões, e coordenação de isolamento em linhas de transmissão de energia elétrica.

Ficha Catalográfica

Rodrigues, Thiago Baptista

Seleção de variáveis e classificação de padrões por redes neurais como auxílio ao diagnóstico de doença cardíaca / Thiago Baptista Rodrigues; orientador: Carlos Kubrusly; co-orientador: José Leonardo Ribeiro Macrini. – 2006.
83 f. : il. ; 30 cm

Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.
Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Informação mútua. 3. Seleção de variáveis. 4. Redes neurais. 5. Classificação. 6. Doença cardíaca – Diagnóstico. I. Kubrusly, Carlos. II. Macrini, José Leonardo Ribeiro. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Agradecimentos

Existe um número muito grande de variáveis, representadas por pessoas e instituições, que deram sua fundamental contribuição para formar a engrenagem que permitiu a elaboração deste trabalho. Certamente não conseguirei enumerar todas as peças dessa engrenagem, tamanha a quantidade, mas prometo tentar.

Ao programa de Mestrado de Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio pela confiança em mim depositada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro concedido.

Aos meus orientadores, o Professor Dr. Carlos Kubrusly e o Professor Dr. José Leonardo Ribeiro Macrini, pela paciência, disponibilidade e conhecimento compartilhado.

Aos Professores integrantes da Banca examinadora, pela contribuição crítica fundamental ao enriquecimento do trabalho.

Ao Centro de Pesquisas de Energia Elétrica (CEPEL), em especial ao Cabral, pelo apoio concedido ao longo do curso.

A Deus, pela força concedida sempre e, principalmente, nos momentos mais difíceis, nos quais nunca me faltou.

Aos meus pais, Rosaly Baptista Rodrigues e Francisco Rodrigues, pela educação que me proporcionaram, pelo empenho para investir em meus estudos, e pelo apoio psicológico nas horas em que mais precisei.

À minha avó, Arinda Coelho Baptista, da qual tenho muita saudade, por tudo que me ensinou, pelos conselhos que me deu, pelo exemplo de ser humano que foi, e pelo auxílio espiritual que me concede sempre.

Por fim, agradeço à minha noiva Rafaela, pelo incentivo e ajuda recebidos, que permitiram me projetar na carreira que escolhi, e também por se fazer sempre presente, mesmo sendo às vezes à distância, o que certamente me possibilitou vencer obstáculos que surgiram pelo caminho com maior facilidade.

Resumo

Rodrigues, Thiago Baptista; Kubrusly, Carlos (Orientador). **Seleção de Variáveis e Classificação de Padrões por Redes Neurais como Auxílio ao Diagnóstico de Doença Cardíaca.** Rio de Janeiro, 2006. 83p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio.

Esta dissertação propõe uma metodologia, baseada em procedimentos quantitativos, para auxiliar o diagnóstico de indivíduos portadores de doença cardíaca. A metodologia proposta foi implementada e analisada em um grupo de indivíduos do banco de dados público intitulado "Heart Disease Database" (Base de Dados pública de Doença Cardíaca) (Aha, atualizado em 2001), diagnosticados nas cidades de Cleveland e Long Beach, nos Estados Unidos. Os resultados obtidos neste estudo foram comparados aos resultados de outros autores encontrados na literatura, de forma a se ter uma medida da qualidade dos resultados aqui obtidos. Foram utilizadas também outras técnicas de classificação de padrões conhecidas na literatura, denominadas "Análise Discriminante" e "Algoritmo C4.5", de forma a estabelecer comparações com os resultados obtidos nesta dissertação utilizando "Redes Neurais", e aplicar a metodologia sugerida na divisão dos conjuntos de treinamento/generalização. Os resultados obtidos foram satisfatórios. Um percentual de acerto médio de 91,0 % foi atingido, enquanto que outros resultados de estudos usando a mesma base de dados alcançaram percentuais de acerto médio de 83,0 % (Ho & Chou, 2001) e 83,5 % (Hu, Li, Cai & Xu, 2004). O desempenho da Rede Neural também foi melhor quando comparado ao da Análise Discriminante e do Algoritmo C4.5. A metodologia de divisão dos conjuntos de treinamento/generalização sugerida nesta dissertação promoveu melhorias em todas as três técnicas de classificação de padrões utilizadas. Acredita-se que os resultados obtidos poderão auxiliar as condutas médicas em relação ao diagnóstico de doença cardíaca, podendo, portanto, vir a ser úteis na prevenção e/ou tratamento de doenças cardíacas.

Palavras-chave

Informação Mútua; Seleção de Variáveis; Redes Neurais; Classificação; Doença Cardíaca; Diagnóstico.

Abstract

Rodrigues, Thiago Baptista; Kubrusly, Carlos (Advisor). **Selection of Variables and Pattern Classification by Neural Networks as Help to the Diagnostic of Heart Disease.** Rio de Janeiro, 2006. 83p. MSc. Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio.

This dissertation proposes a methodology, established in quantitative procedures, to assist the diagnostic of individuals with heart disease. The proposed methodology was implemented and analyzed in a group of individuals of the public database called “Heart Disease Database” (Aha, current in 2001), diagnosed in the cities of Cleveland and Long Beach, in the United States. The results gotten in this study had been compared with the results of other authors found in literature to have a measure of the quality of the results gotten here. Others techniques of classification of standards known in literature had also been used, called “Discriminate Analysis” and “C4.5 Algorithm”, to establish comparisons with the results gotten in this dissertation using “Neural Networks”, and to apply the methodology suggested in the division of the sets of training/generalization. The gotten results were satisfactory. A percentage of average rightness of 91.0 % was reached, whereas other results of studies using the same database had reached percentages of average rightness of 83.0 % (Ho & Chou, 2001) and 83.5 % (Hu, Li, Cai & Xu, 2004). The performance of the Neural Network was also better when compared with Discriminate Analysis and C4.5 Algorithm. The methodology of division of the sets of training/generalization suggested in this dissertation promoted improvements in all the three used techniques of classification of standards. It’s believable that the gotten results will be able to assist the medical behaviors in relation to the diagnostic of heart disease, becoming useful in the prevention and/or treatment of heart diseases.

Keywords

Mutual Information; Selection of Variables; Neural Networks; Classification; Heart Disease; Diagnostic.

Sumário

Lista de Figuras

Lista de Tabelas