

**SANDRO MORETTI CORREIA DE ALMEIDA**

**MODELO GAC-BRASIL - ALGORITMO GENÉTICO APLICADO À  
SIMILARIDADE PARA RACIOCÍNIO BASEADO EM CASOS - ESTUDO  
DE CASO EM DIAGNÓSTICO DE CARDIOPATIA ISQUÊMICA**

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Gestão do Conhecimento e da Tecnologia da Informação, da Universidade Católica de Brasília, como requisito parcial para obtenção do Título de Mestre em Gestão do Conhecimento e da Tecnologia da Informação.

Orientadora:  
Prof.<sup>a</sup> Dr.<sup>a</sup> Lourdes Mattos Brasil

Co-orientadores:  
Prof. Dr. Edilson Ferneda  
Prof. Dr. Hervaldo Sampaio Carvalho

BRASÍLIA  
2006

A447m Almeida, Sandro Moretti Correia de.

Modelo GAC-BRASIL: algoritmo genético aplicado à similaridade para raciocínio baseado em casos: estudo de caso em diagnóstico de cardiopatia isquêmica / Sandro Moretti Correia de Almeida – 2006.

89 f.: il. ; 30 cm

Dissertação (mestrado) – Universidade Católica de Brasília, 2006.

Orientação: Lourdes Mattos Brasil

Co-orientação: Edilson Ferneda; Hervaldo Sampaio de Carvalho

1. Algoritmos genéticos. 2. Raciocínio. 3. Aprendizado do computador.  
4. Diagnóstico. 5. Cardiopatia . I. Brasil, Lourdes Mattos, orientadora.  
II. Ferneda, Edilson, co-orientador. III. Carvalho, Hervaldo Sampaio de,  
co-orientador. IV. Título

CDU 616-071:004

Ficha elaborada pela Coordenação de Processamento do Acervo do SIBI – UCB.

# TERMO DE APROVAÇÃO

MODELO GAC-BRASIL - ALGORITMO GENÉTICO APLICADO À SIMILARIDADE  
PARA RACIOCÍNIO BASEADO EM CASOS - ESTUDO DE CASO EM DIAGNÓSTICO  
DE CARDIOPATIA ISQUÊMICA

Dissertação defendida e aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Gestão do Conhecimento e Tecnologia da Informação, defendida e aprovada, em 01 de dezembro de 2006, pela banca examinadora constituída por:

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Lourdes Mattos Brasil  
Orientadora

---

Prof. Dr. Edilson Fereda

---

Prof. Dr. Hervaldo Sampaio Carvalho

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Germana Menezes da Nóbrega

---

Prof. Dr. Heitor Silvério Lopes

## DEDICATÓRIA

*Este trabalho é dedicado à memória de Anderson e Wingston,  
irmãos que marcaram duas das primeiras fases da minha vida,  
a formação da pessoa e do profissional,  
e cuja lembrança e saudade, amadureceram meu coração para as fases seguintes.*

## AGRADECIMENTOS

*Agradeço em primeiro lugar a Deus, que ilumina todos os meus caminhos.*

*Aos meus pais, Messias e Zenilda, que tornaram este trabalho possível através do seu amor e do sacrifício de prazeres próprios em prol da educação dos filhos.*

*À minha esposa Márcia, pelo amor e apoio incondicional neste e em todos os projetos da nossa vida.*

*Aos meus filhos, Anderson e Marina, luzes da minha vida e razões da minha luta.*

*Aos meus irmãos, Wesley e Walisson, dois exemplos de caráter e atitude.*

*Agradeço aos meus orientadores, Profa. Dra. Lourdes Brasil, Prof. Dr. Edilson Ferneda e Prof. Dr. Hervaldo Carvalho, pela disciplina exigida, cobrança carinhosa, direcionamento seguro, contribuições importantes, colocações pertinentes e observações pragmáticas. Este trabalho é de vocês também.*

*Agradecimento especial à colega Renata Silva por ter me confiado a continuidade de sua pesquisa, e pelo precioso tempo de descanso dedicado a me ajudar.*

*Agradeço também à amiga Cristiane Ramos pelo incentivo a fazer este curso, aos professores Gentil Lucena e Cláudio Chauke pelo apoio na primeira batalha para estar aqui, aos demais professores do programa pela grandeza em compartilhar conhecimento e às sempre carinhosas e prestativas Janina e Georgiane, pela ajuda ao longo do mestrado.*

*Por último, mas não menos importante, agradeço aos demais familiares, colegas e ex-colegas de trabalho e de mestrado, meus alunos, amigos da minha paróquia e companheiros de farrã, que com dicas, críticas e palavras de incentivo, ou simplesmente ouvindo pacientemente, também ajudaram a construir este trabalho.*

## EPÍGRAFE

*“Nunca deixe que lhe digam  
que não vale a pena  
acreditar no sonho que se tem,  
ou que seus planos nunca vão dar certo,  
ou que você nunca vai ser alguém...  
Se você quiser alguém em quem confiar,  
confie em si mesmo.  
Quem acredita sempre alcança.”*

*Trechos de “Mais ma vez” por Renato Russo*

## RESUMO

O objetivo do Raciocínio Baseado em Casos (RBC) é a reutilização de soluções conhecidas no contexto de um problema novo, de solução ainda desconhecida, através da busca de casos semelhantes. Esta determinação de exemplos de casos similares é um dos problemas centrais desta técnica da Inteligência Artificial. O módulo RBC do Projeto Inteligência Artificial Aplicada na Modelagem e Implementação de um Consultório Virtual (IACVIRTUAL) da Universidade Católica de Brasília utiliza uma função de similaridade global que computa a soma ponderada dos resultados das funções de similaridade local de cada característica do caso, sendo que o peso atribuído para cada característica foi definido por um especialista do campo da Cardiopatia Isquêmica. Esta forma de atribuição, embora bastante utilizada, é baseada somente na experiência e conhecimento de uma ou mais pessoas, sendo, portanto, subjetiva e não fundamentada em bases estatísticas. Este trabalho define um modelo de Algoritmo Genético aplicado à Similaridade para Raciocínio Baseado em Casos, que foi denominado GAC-BRASIL a partir do inglês *Genetic Algorithm applied to Case-Based Reasoning SimiLarity*. O Algoritmo Genético, técnica de Aprendizado de Máquina baseada na Computação Evolucionária, cobre o espaço de busca combinando a sobrevivência de estruturas de valores melhor adequadas à solução e a troca de informações entre estas de uma maneira estruturada e ao mesmo tempo aleatória, num mecanismo de otimização que imita a genética e a seleção natural. Será apresentado um estudo de caso validado por um especialista médico da área de Cardiologia, onde o modelo GAC-BRASIL foi utilizado para a criação de um módulo que determina os pesos ideais de cada característica no cálculo da função de similaridade global do módulo RBC no contexto do Projeto IACVIRTUAL com casos reais de Cardiopatia Isquêmica.

**Palavras-chave:** Algoritmo Genético, Raciocínio Baseado em Casos, Cálculo de Similaridade, Aprendizagem de Máquina, Diagnóstico Médico, Cardiopatia Isquêmica.

## ABSTRACT

The goal of Case-Based Reasoning (CBR) is to reuse known solutions to solve new problems by searching for similar cases in a base of cases. Similarity establishing criteria is one of the major issues of this Artificial Intelligence technique. The CBR module of Artificial Intelligence Applied in the Modeling and Implementation of a Virtual Medical Office Project (IACVIRTUAL) uses a global similarity function that computes the weighed sum of the local similarity functions previous results for each symptom or characteristic between the cases. The weighs that indicated the relevance used by this function have been indicated by an Ischemic Cardiopathy expert. Although this method is largely applied, this definition totally relies on the skill and medical knowledge of one or more experts, therefore is subjective and non statistics oriented. Genetic Algorithm is a Learning Machine method based on Evolutionary Computation that covers the search space by combining the best fitness structures survival and the information exchange between them, in a structured and random way as well. Such search mechanism emulates the natural genetic and the natural selection. This work defines a model of Genetic Algorithm applied to Case-Based ReAsoning SimiLarity named GAC-BRASIL, and it will present a case study validated by a Cardiology medical expert. In this validation with real Ischemic Cardiopathy cases, the GAC-BRASIL model has been used to create a system to find the most appropriated weighs for each symptom or characteristic to be used on the global similarity function of the CBR module of IACVIRTUAL Project.

**Keywords:** Genetic Algorithm, Case-Based Reasoning, Similarity Calculation, Learning Machine, Medical Diagnosis, Ischemic Cardiopathy.



# SUMÁRIO

<b>Lista de Abreviaturas e Símbolos .....</b>	<b>12</b>
<b>1. Introdução .....</b>	<b>14</b>
1.1. Tema .....	15
1.2. Revisão de literatura .....	15
1.3. Justificativa .....	19
1.4. Formulação do problema .....	20
1.5. Objetivos .....	21
1.6. Organização do trabalho .....	21
<b>2. Referencial Teórico.....</b>	<b>22</b>
2.1. Gestão do conhecimento médico .....	23
2.1.1. Gestão do conhecimento.....	23
2.1.2. Conhecimento médico .....	24
2.2. Raciocínio baseado em casos e o domínio médico .....	25
2.3. Cálculo de similaridade.....	29
2.4. Aprendizado de máquina .....	30
2.4.1. Aprendizado por árvores de decisão.....	31
2.4.2. Redes neurais artificiais.....	32
2.4.3. Redes bayesianas .....	32
2.4.4. Aprendizado de máquina baseado em genética.....	33
2.5. Algoritmos genéticos .....	35
2.5.1. Computação evolucionária .....	35
2.5.2. Algoritmos genéticos.....	35
2.5.3. Problema de otimização.....	36
2.5.4. Terminologia .....	37
2.5.5. Operadores genéticos.....	41
2.5.6. Utilização de números reais.....	44
2.6. Projeto IACVIRTUAL.....	46
<b>3. Metodologia .....</b>	<b>49</b>
3.1. Classificação da pesquisa.....	50
3.2. Suposições e hipóteses .....	50
3.3. Coleta e análise de dados .....	50
3.4. Delimitação do estudo.....	51
3.5. Resultados esperados .....	51
3.6. O modelo GAC-BRASIL.....	52
3.6.1. Parâmetros e definições.....	52
3.6.2. A função <i>fitness</i> .....	55
<b>4. Resultados.....</b>	<b>58</b>
4.1. Implementação de interface para o módulo RBC .....	59
4.1.1. Problema corrente.....	60
4.1.2. Similaridade local.....	60
4.1.3. Similaridade global.....	62
4.2. Implementação do módulo AG.....	63

4.2.1. Preparação dos dados para a função <i>fitness</i> .....	64
4.2.2. Configuração dos parâmetros do AG .....	65
4.3. Representação por dígitos binários x números reais .....	69
4.4. Definição de pesos por um grupo de especialistas médicos .....	71
4.5. Utilização dos pesos no módulo RBC .....	73
4.5.1. Escolha do grupo de pesos.....	73
4.5.2. Apresentação dos resultados do RBC.....	75
4.6. Simulações do RBC para avaliação dos resultados do AG.....	76
4.6.1. Primeiras simulações e ajuste na base de dados .....	77
4.6.2. Simulações de RBC .....	78
<b>5. Conclusões e Trabalhos Futuros .....</b>	<b>80</b>
5.1. Conclusões .....	81
5.2. Trabalhos futuros .....	82
<b>Referências .....</b>	<b>84</b>

## Lista de Figuras

Figura 1 - Ciclo RBC (AAMODT E PLAZA, 1994) .....	28
Figura 2 - <i>Crossover</i> de um ponto .....	42
Figura 3 - Mutação em codificação binária .....	43
Figura 4 - Projeto IACVIRTUAL .....	47
Figura 5 - Fluxograma básico do GAC-BRASIL .....	53
Figura 6 - Fluxograma da função de aptidão ( <i>fitness</i> ) .....	56
Figura 7 - Tela de informações do problema corrente .....	60
Figura 8 - Tela de parametrização do AG .....	66
Figura 9 - Gráfico de evolução do <i>fitness</i> em AG com números reais .....	70
Figura 10 - Gráfico de evolução do <i>fitness</i> em AG com dígitos binários .....	70
Figura 11 - Formulário para indicação de pesos por especialistas .....	72
Figura 12 - Gráfico comparativo entre pesos atribuídos pelos especialistas e pelo AG .....	74
Figura 13 - Tela de escolha do conjunto de pesos .....	75
Figura 14 - Tela de apresentação dos resultados do RBC .....	76

## Lista de Tabelas

Tabela 1 - Aplicações de AM Baseadas em Genética (GOLDBERG, 1989).....	34
Tabela 2 - Terminologia dos AG x Biologia (LOPES, 1996; CORRÊA, 2000).....	38
Tabela 3 - Características e sintomas dos casos .....	61
Tabela 4 - Funções de similaridade local .....	62
Tabela 5 - Pesos do modelo atual .....	63
Tabela 6 - Composição do grupo de “problemas correntes” da função <i>fitness</i> .....	65
Tabela 7 - Composição do grupo de “casos da base” da função <i>fitness</i> .....	65
Tabela 8 - Comparação roleta x <i>ranking</i> .....	68
Tabela 9 - Valores médios dos genes das melhores soluções do AG.....	71
Tabela 10 - Média dos pesos atribuídos pelo grupo de especialistas .....	73
Tabela 11 - Comparação dos pesos atribuídos pelos especialistas e pelo AG .....	74
Tabela 12 - Média de acertos (mesmo diagnóstico) por tipo de peso .....	78

## Lista de Abreviaturas e Símbolos

AE - Algoritmos Evolucionários

AM - Aprendizagem de Máquina

AMBG - Aprendizado de Máquina Baseado em Genética

BDTD - Banco de Dados de Teses e Dissertações do IBICT

CBRG - *Case-Based Repair Generation*

CCT - Centro de Ciência e Tecnologia da Universidade Católica de Brasília

CE - Computação Evolucionária

CRISP-DM - *Cross-industry Standard Process for Data Mining*

CS1 - *Cognitive System Level One*

GAC-BRASIL - Algoritmo Genético aplicado à Similaridade para Raciocínio Baseado em Casos / *Genetic Algorithm applied to Case-Based Reasoning SIMilarity*

HUB - Hospital Universitário da Universidade de Brasília

IA - Inteligência Artificial

IACVIRTUAL - Projeto Inteligência Artificial Aplicada na Modelagem e Implementação de um Consultório Virtual

IAM - Infarto Agudo do Miocárdio

IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia

MGCTI - Mestrado em Gestão do Conhecimento e da Tecnologia da Informação da Universidade Católica de Brasília

OG - Operadores Genéticos

RBC - Raciocínio Baseado em Casos

RNA - Redes Neurais Artificiais

RV - Realidade Virtual

SC - Sistema Cognitivo

SEH - Sistema Especialista Híbrido

STI - Sistema Tutor Inteligente

SWAM - Sistema Web para Análise de Mamografia Digital

TI - Tecnologia da Informação

UCB - Universidade Católica de Brasília

UERJ - Universidade Estadual do Rio de Janeiro

UnB - Universidade de Brasília

UNICAMP - Universidade Estadual de Campinas

USP - Universidade de São Paulo

# **1.Introdução**

---

## 1.1. Tema

O tema onde este trabalho se insere é a utilização de Algoritmo Genético (AG), uma técnica de Aprendizagem de Máquina (AM) baseada em Computação Evolucionária (CE), para otimização do cálculo de similaridade em um modelo de Raciocínio Baseado em Casos (RBC) de suporte ao Diagnóstico Médico na área de Cardiologia.

Mais especificamente, será proposto um modelo de AG para determinação dos pesos de características a serem considerados no cálculo da similaridade global em um sistema de RBC. Este modelo, chamado GAC-BRASIL (Algoritmo Genético aplicado à Similaridade para Raciocínio Baseado em Casos / *Genetic Algorithm applied to Case-Based Reasoning Similarity*), será validado em uma base de casos reais com diagnósticos conhecidos de Cardiopatia Isquêmica.

## 1.2. Revisão de literatura

Buscando soluções publicadas para o problema da definição de pesos de características para RBC, foram feitas pesquisas em bibliografias especializadas no que diz respeito a livros e às diversas bases de teses, monografias e artigos: UCB, UNICAMP, USP, IBICT, UERJ/COPPE e *Google Scholar*.

A tese de doutorado de Hall (1999) endereça o problema de tarefas de classificação em AM, sua pesquisa é baseada na hipótese de que bons conjuntos de características contêm características altamente relacionadas com a classe, ainda que não relacionadas entre si. O trabalho de Wettschereck e Aha (1995) apresenta um *framework* de cinco dimensões para resolver o problema de características irrelevantes e que causam distorções em resultados da função “vizinho mais próximo”. Estas pesquisas mostram que a seleção do conjunto de características adequadas para a tarefa de classificação é um ponto chave em AM.



Trabalhos recentes como o de Rodrigues (2004) do programa de Engenharia Civil da COPPE/UERJ, que utiliza AG para otimização de configurações dutos de escoamento de petróleo em plataformas marítimas; o de Tsunoda (2004) da área de concentração em Engenharia Biomédica do programa de Engenharia Elétrica e Informática Industrial do CEFET/PR, que utiliza conceitos de Computação Evolucionária para conceber sistemas adequados à classificação de proteínas; e o de Silva (1999) do INPE/MCT, que utiliza AG para resolver o problema de correspondência em estereoscopia; são casos em que Algoritmos Evolucionários endereçaram com sucesso problemas de otimização de diferentes áreas de estudo.

O Banco de Dados de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), quando pesquisado com a palavra-chave “Algoritmo”, retorna 682 trabalhos. Já com o termo “Algoritmo Genético” foram encontrados 65 trabalhos. Quando acrescido o termo “Similaridade” temos apenas o já citado trabalho de Silva (1999) e o de Drummond (2003), que usa similaridade para tratamento de consistência e cobertura em sistemas de regras difusas. O termo “RBC” sozinho retorna 21 trabalhos, mas, quando usado em conjunto com “Algoritmo Genético”, não há nenhum retorno.

Na base da Universidade de São Paulo (USP), que tem mais de 4.300 trabalhos entre teses e dissertações, os mesmos critérios de pesquisa retornaram sete trabalhos para “Algoritmo Genético” e cinco para “RBC”. Nenhum trabalho foi encontrado com as palavras-chave “Algoritmo Genético” e “RBC”, tampouco com “Algoritmo Genético” e “Similaridade”.

A base da Universidade Estadual de Campinas (UNICAMP) tem 14.377 documentos e 7.511 teses. Apenas 49 documentos têm referência registrada a “Algoritmo Genético” e, destes, 18 também se relacionam com o termo “Similaridade”. Um total de seis trabalhos cita o termo “Raciocínio Baseado em Casos”, dos quais dois contêm a combinação de “Algoritmo

Genético” e “RBC”. Contudo, um deles, a tese de doutorado de Silva (2001), apenas menciona o termo RBC quando cita a comparação feita por Hunt e Fellows (1996 *apud* SILVA, 2001) desta técnica com seu sistema imunológico artificial, não tendo relação entre RBC e AG. Assim o resultado desta procura reduziu-se à tese de doutorado em Engenharia Mecânica de Mendes (2001), que aborda o uso de RBC em projetos de perfuração de poços de petróleo, utilizando experiências anteriores, principalmente referentes à geologia, com o objetivo de reduzir custos. No entanto, enquanto o presente trabalho pretende utilizar AG como parte do mecanismo de busca, Mendes (2001) utiliza AG na etapa de adaptação, implementando uma técnica para determinar a relevância dos casos já recuperados da base.

Dentre os trabalhos internacionais nesta área, e mais próximos do problema abordado nesta pesquisa, temos o de Demiroz e Guvenir (1996), que propõem um AG para determinar os pesos de características para um algoritmo de classificação utilizando a função “vizinho mais próximo”. A contribuição deste artigo está na proposta de um algoritmo com uma variação do operador genético de *crossover*, que foi chamada de *continuous uniform crossover*, que, em sua pesquisa, apresentou melhores resultados que algoritmos com os operadores *one-point crossover*, *two-point crossover* e *uniform crossover*. Esta pesquisa, ao contrário do presente trabalho, focou basicamente na comparação entre os métodos de *crossover*, inclusive não utilizando mutação em seus algoritmos.

O artigo de Wang e Ishii (1997), embora também não trate de RBC, propõe um AG para aprendizado de pesos em uma função de similaridade linear e outra não-linear. Em relação a um AG básico, a proposta tem as seguintes características: (i) utilização de números reais para representar os genes dos indivíduos; (ii) todos os membros se reproduzem, sem haver seleção; (iii) há o que pode ser considerada uma variação do elitismo onde, após a reprodução, é calculada a aptidão do filho, e se esta for menor que a de um dos pais, este pai é quem será

copiado para a próxima geração; (iv) a mutação é utilizada para resolver o problema de indivíduos iguais e é aplicada na geração resultante até não haver vetores iguais.

Um trabalho recente e que explicitamente associa RBC e AG é o de Beddoe e Petrovic (2004). A proposta apresentada é a evolução de sua técnica, denominada *Case-Based Repair Generation* (CBRG) que usa RBC para resolver o problema de plantões (turnos de trabalho) de enfermeiras (BEDDOE E PETROVIC, 2003). O CBRG pode ser visto como um sistema classificador que determina o tipo e os parâmetros de reparo para uma dada violação. Neste trabalho de 2004, os autores propõem a utilização de AG para determinar os pesos da função de distância “vizinho mais próximo” com o objetivo de reduzir o número de características consideradas no problema, visando aumentar a qualidade e eficiência do CBRG.

O já citado artigo de Wettschereck e Aha (1995), além de associar RBC e AG, assim como o presente trabalho, também trata de similaridade, mais especificamente dos pesos de características na função de similaridade “vizinho mais próximo”. Os autores afirmam que esta é a função mais utilizada no cálculo de similaridade em RBC, mas a consideram muito sensível a características irrelevantes, que podem causar distorções no resultado. Contudo, ao contrário do presente trabalho, que pretende utilizar AG para calibrar os pesos das características, a proposta de Wettschereck e Aha (1995) é detectar e eliminar as características consideradas irrelevantes.

Da associação entre AG, RBC e Diagnóstico na área de Saúde, temos a tese de Lopes (1996). Neste trabalho é utilizado o AG para otimizar o peso de cada atributo para o diagnóstico de 12 patologias no contexto de raciocínio por analogia, do qual o RBC é um caso particular. Este trabalho, por abordar o problema de representação do conhecimento em problemas tratados por Inteligência Artificial (IA), utilizar um protótipo de sistema baseado em raciocínio por analogia no tratamento de Dor Torácica (DoToR) e implementar com sucesso um AG para otimização de pesos de atributos no processo de recuperação de casos

(sistema GADoToR), é o trabalho que mais se assemelha ao tema tratado na presente pesquisa.

Pode-se observar, pelo material pesquisado, que AG é um método robusto e bastante utilizado para calibragem de pesos em funções ponderadas, inclusive envolvendo similaridade. Contudo, ainda nota-se bastante espaço para pesquisa na combinação desta técnica com o processo de recuperação de casos em RBC e, especificamente, na área da saúde.

### **1.3. Justificativa**

O ser humano aprende principalmente por experiência. Dentro do campo da IA, o RBC imita esta característica ao tentar resolver uma situação problema corrente com base em uma busca de casos similares previamente armazenados em uma base de casos. Esta busca de casos similares é um dos problemas centrais do RBC (WANGENHEIM E WANGENHEIM, 2003).

O Projeto IACVIRTUAL (ALMEIDA et al., 2003a, 2003b, 2004a, 2004b) é um sistema baseado em conhecimento que se propõe a simular um consultório virtual via WEB, concebido para atender *(i)* pacientes interessados em acompanhar seu histórico médico; *(ii)* especialistas interessados num ambiente de apoio à decisão em diagnóstico e tratamento de seus pacientes e *(iii)* estudantes interessados na aprendizagem por meio do acompanhamento dos casos médicos disponíveis.

O módulo RBC do Projeto IACVIRTUAL utiliza funções ponderadas no cálculo de similaridade local e global para subsidiar a busca de casos similares na base. Atualmente, os pesos de cada característica no cálculo da similaridade foram atribuídos por um especialista médico e, em 2005, Silva (2005a, p.90) destacava que uma “investigação importante a ser

realizada no projeto é a aprendizagem de máquina, apoiando o Engenheiro do Conhecimento e o especialista nesta tarefa de definição de pesos”.

Como o sucesso do Diagnóstico por RBC depende fundamentalmente da seleção dos casos mais similares, busca-se uma maneira de otimizar esta definição dos pesos.

Segundo Mitchel (1997), AM endereça a questão de como construir programas de computador que melhoram seu desempenho em uma tarefa através da experiência. Dentre os métodos de AM estão os AG. AG são métodos de otimização e busca inspirados nos mecanismos de evolução dos seres vivos, onde otimização consiste em tentar várias soluções e utilizar a informação obtida neste processo de forma a encontrar soluções cada vez melhores (LACERDA E CARVALHO, 1999).

Este trabalho, uma pesquisa que é fundamentalmente motivada pela necessidade de resolver problemas concretos, tendo, portanto, finalidade prática, visa investigar o problema da atribuição de pesos de características do atual modelo RBC do IACVIRTUAL, propondo e aplicando um modelo baseado em AG que promova um suporte a esta tarefa, auxiliando o Engenheiro do Conhecimento e o especialista médico.

#### **1.4. Formulação do problema**

Como citado no tópico anterior, o atual modelo RBC do Projeto IACVIRTUAL tem a atribuição de pesos das características de maneira totalmente subjetiva por parte do especialista. No entanto, a área de saúde precisa de métodos rápidos e eficientes. O IACVIRTUAL não pretende substituir o especialista médico e sim ser um sistema de suporte à decisão.

Assim, a idéia é manter a atribuição de pesos pelo especialista, mas apresentar uma forma de suporte a esta decisão com o cálculo automático dos pesos das características.

## 1.5. Objetivos

O objetivo geral deste trabalho é a concepção e a implementação de um método de aprendizagem de máquina para auxiliar o especialista médico na atribuição de pesos de características de casos no cálculo da similaridade global de um sistema baseado em RBC.

Os objetivos específicos são:

- Propor um modelo para utilização de AG na calibragem de pesos para RBC;
- Implementar um módulo utilizando o modelo proposto para o cálculo automático de pesos das características dos casos do modelo RBC do Projeto IACVIRTUAL;
- Otimizar a ferramenta de busca e recuperação de casos do módulo RBC do Projeto IACVIRTUAL;
- Validar o modelo proposto numa base de casos reais na área de Cardiologia, contendo diagnósticos de Cardiopatia Isquêmica;
- Comparar os resultados da melhor solução do AG proposto com os resultados obtidos com a definição de pesos pelo especialista médico.

## 1.6. Organização do trabalho

Este trabalho está organizado em cinco capítulos, incluindo este capítulo.

No capítulo dois, é apresentada uma visão geral do referencial teórico deste trabalho, visando uma compreensão de RBC e de AG. A metodologia para utilização do AG no modelo RBC do Projeto IACVIRTUAL é apresentada no capítulo três.

No capítulo quatro são descritos os resultados obtidos, assim como as técnicas utilizadas para calibragem do AG.

Por fim, as conclusões e sugestões de trabalhos futuros são apresentadas no quinto capítulo.

## **2.Referencial Teórico**

---

## **2.1. Gestão do conhecimento médico**

### **2.1.1. Gestão do conhecimento**

A gestão do conhecimento pode ser vista como uma coleção de processos que governa a criação, disseminação e utilização do conhecimento para atingir plenamente os objetivos da organização (TEIXEIRA FILHO, 2001).

A gestão do conhecimento também pode ser definida como o conjunto de atividades que busca desenvolver e controlar todo tipo de conhecimento em uma organização, visando sua utilização na consecução de seus objetivos. Esse conjunto de atividades deve ter como principal meta apoiar o processo decisório, em todos os níveis. Para tanto, é preciso estabelecer políticas, procedimentos e tecnologias que sejam capazes de coletar, distribuir e utilizar efetivamente o conhecimento, representando fator de mudança no comportamento organizacional (MORESI, 2001).

O conhecimento organizacional pode ser classificado em dois tipos (NONAKA E TAKEUCHI, 1997). Um é o conhecimento explícito, que pode ser articulado na linguagem formal, sobretudo em afirmações gramaticais, expressões matemáticas, especificações, manuais e assim por diante. Esse tipo de conhecimento pode ser então transmitido, formal e facilmente, entre os indivíduos.

O segundo tipo, o conhecimento tácito, é difícil de ser articulado na linguagem formal. É o conhecimento pessoal incorporado à experiência individual e envolve fatores intangíveis como, por exemplo, crenças pessoais, perspectivas e sistemas de valor.

Stollenwerk (2001) constatou a importância do processo criação do conhecimento, comum a todos os modelos por ela estudados, e, dentro desse processo, a dimensão aprendizagem organizacional foi percebida como essencial para operacionalização do modelo.



Concluindo, a atividade fundamental da Gestão do Conhecimento numa organização é mapear dinamicamente os conhecimentos, experiências e práticas, registrando e assegurando que o conhecimento esteja disponível nos locais de tomada de decisão e mantendo-o sempre atualizado.

### **2.1.2. Conhecimento médico**

A profissão médica tem sofrido significativas mudanças devido ao desenvolvimento e aos avanços científicos das áreas concernentes. Assim como na grande maioria das profissões atuais, profissionais da área médica devem constantemente se atualizar. A geração e comunicação de conhecimento por parte de médicos é, hoje, um importante elemento do sucesso e um desafio profissional (PORTAL MÉDICO, 2005 *apud* SILVA, 2005a).

Por outro lado, segundo Davenport e Prusak (1998), seres humanos aprendem melhor com histórias. A pesquisa de Schank (1982) e o trabalho de sua aluna Kolodner (1993) mostram que o conhecimento é comunicado com mais eficácia por meio de uma narrativa convincente, elegante e apaixonada. A utilização de narrativas é uma das melhores maneiras de ensinar e aprender coisas complexas. Na maioria das vezes, é possível estruturar histórias de forma a transmitir significado sem perda substancial de seu poder de comunicação.

Davenport e Prusak (1998) sugerem que o RBC é uma forma de combinar em computador o poder da narrativa com a codificação do conhecimento. A tecnologia envolve a extração do conhecimento de uma série de narrativas, ou casos, sobre a área de problema. RBC, diferentemente dos sistemas especializados, que exigem regras bem estruturadas e isentas de duplicidade, permite que as estruturas de casos possam refletir um pensamento.

Durante o processo de diagnóstico, o profissional de saúde realiza uma série de inferências sobre a natureza das disfunções do corpo. Estas inferências são derivadas das observações existentes, ou seja, dados consistentes sobre a história do paciente, sinais, sintomas, testes de rotinas, respostas a várias manipulações, o tempo de curso de alguns

eventos, os conhecimentos clínicos, fisiológicos, bioquímicos, anatômicos e patológicos, sobre casos semelhantes e a sua biologia subjacente, a experiência prévia do médico em realizar diagnósticos do mesmo tipo, bem como o senso comum e a intuição (KASSIRER E GORRY, 1978).

Os profissionais de saúde utilizam amplamente suas experiências passadas no processo de reunir e interpretar informações. Estas experiências são essenciais, pois reduzem a necessidade de se ter um número considerável de questões desnecessárias, de ordenar testes de diagnósticos supérfluos e de tornar a tarefa de informação manejável e eficiente (KASSIRER E GORRY, 1978).

## **2.2. Raciocínio baseado em casos e o domínio médico**

A cardiopatia isquêmica, tomada como domínio de aplicação deste trabalho, é uma doença cardíaca decorrente da diminuição significativa do fluxo coronariano devido à formação de placas de gordura (ateroma) em determinados segmentos das artérias coronarianas e/ou redução do calibre por espasmos (RABELO JR. et al., 1993).

Algumas características dos sintomas devem ser cuidadosamente indagadas com a finalidade de orientarem a probabilidade da presença de angina (ARQUIVOS BRASILEIROS DE CARDIOLOGIA, 2004):

- qualidade (constritiva, aperto, peso, opressão, desconforto, queimação, pontada);
- localização (precordial, retroesternal, ombro, epigástrico, cervical, hemitórax, dorso);
- irradiação (membros superiores direito, esquerdo ou ambos, ombro, mandíbula, pescoço, dorso, região epigástrica);
- duração (segundos, minutos, horas e dias);
- fatores desencadeantes (esforço físico, atividade sexual, posição, alimentação, respiração, componente emocional, espontânea);

- fatores de alívio (repouso, nitrato sublingual, analgésico, alimentação, antiácido, posição e apnéia);
- sintomas associados (sudorese, náusea, vômito, palidez, dispnéia, hemoptise, tosse, pré-síncope e síncope).

Vários são os adjetivos utilizados pelos pacientes na descrição da angina: “sufocamento”, “queimação”, “opressão”, “peso”, entre outros (ARQUIVOS BRASILEIROS DE CARDIOLOGIA, 2004).

Para a elaboração do diagnóstico com grande chance de sucesso para eventos agudos da cardiopatia isquêmica, têm-se como principais elementos (RABELO JR. et al., 1993): *(i)* história clínica do paciente; *(ii)* exame físico e *(iii)* eletrocardiograma. Assim, um posto de saúde, mesmo que pequeno, poderia formular um diagnóstico de evento coronariano agudo, caso tais recursos estejam disponíveis. O que ocorre é que, muitas vezes, mesmo com esses recursos, o diagnóstico não é feito pela falta de conhecimento especializado.

Desta forma, um sistema baseado em experiências como, por exemplo, o RBC, pode ser de extrema validade, pois, a partir da parametrização do ambiente e um grupo de informações iniciais (dados clínicos), o sistema pode sugerir um diagnóstico mais provável através das experiências acumuladas (SILVA, 2005a; ALMEIDA et al., 2006a; ALMEIDA et al., 2006b).

O Projeto IACVIRTUAL, proposto pelo Grupo de Inteligência Artificial da Universidade Católica de Brasília, aborda também este problema utilizando, dentre outras técnicas, um modelo RBC para o diagnóstico de Cardiopatia Isquêmica.

RBC constitui uma tecnologia computacional, do campo da IA, cujas aplicações estão cada vez mais diversificadas. Trata-se de um recurso de modelagem computacional apropriado para soluções e respostas a problemas tais como (MARTINS, 2000):

- Problemas de um médico frente ao seu paciente: Qual o diagnóstico para estes sintomas? Qual tratamento mais indicado?

- Problemas de um operador de alto-forno siderúrgico em atividade de monitoração: Considerando o tipo de liga dentro do alto-forno, devo aumentar a temperatura?
- Problemas de um corretor de imóveis frente a um proprietário: Qual o valor mais apropriado para este imóvel?

RBC é um enfoque para a solução de problemas e para o aprendizado baseado em experiência passada. RBC resolve problemas ao recuperar e adaptar experiências passadas, chamadas de casos, armazenadas em uma base de casos. Um novo problema é resolvido com base na adaptação de soluções de problemas similares já conhecidas. Um grande número de exemplos da vida diária pode ser utilizado para demonstrar como seres humanos utilizam casos conhecidos como uma forma de resolução de problemas de um modo extremamente natural. Por exemplo, ao atender um novo paciente e escutar seus problemas, o médico lembra-se do histórico da doença de um outro paciente devido ao conjunto similar de sintomas, e aplica-lhe um tratamento semelhante ao que administrou ao paciente que apresentou aqueles sintomas similares (WANGENHEIM E WANGENHEIM, 2003).

No RBC problemas são resolvidos exclusivamente no âmbito de um mesmo domínio de aplicação com utilização de exemplos que fazem parte deste domínio (WANGENHEIM E WANGENHEIM, 2003).

A arquitetura RBC pode ser vista, de uma forma mais genérica, pelas etapas definidas na Figura 1, que representa o ciclo de processamento desse sistema (conhecido como R4) (AAMODT E PLAZA, 1994):

- Recuperação ou resgate: recupera-se o caso mais similar ou um conjunto de casos da biblioteca de casos;
- Reutilização: reutilizam-se as informações e o conhecimento que resolveram o problema anteriormente, associando-se sua solução ao contexto do problema atual;

- Revisão: revisa-se (caso seja necessário) e adapta-se a antiga solução para que seja utilizada na nova situação (problema);
- Retenção: é a etapa em que se armazenam todas as partes consideradas úteis do novo problema na biblioteca de casos. Além disso, deve-se decidir de que forma armazenar e como indexar o caso para futuras recuperações.

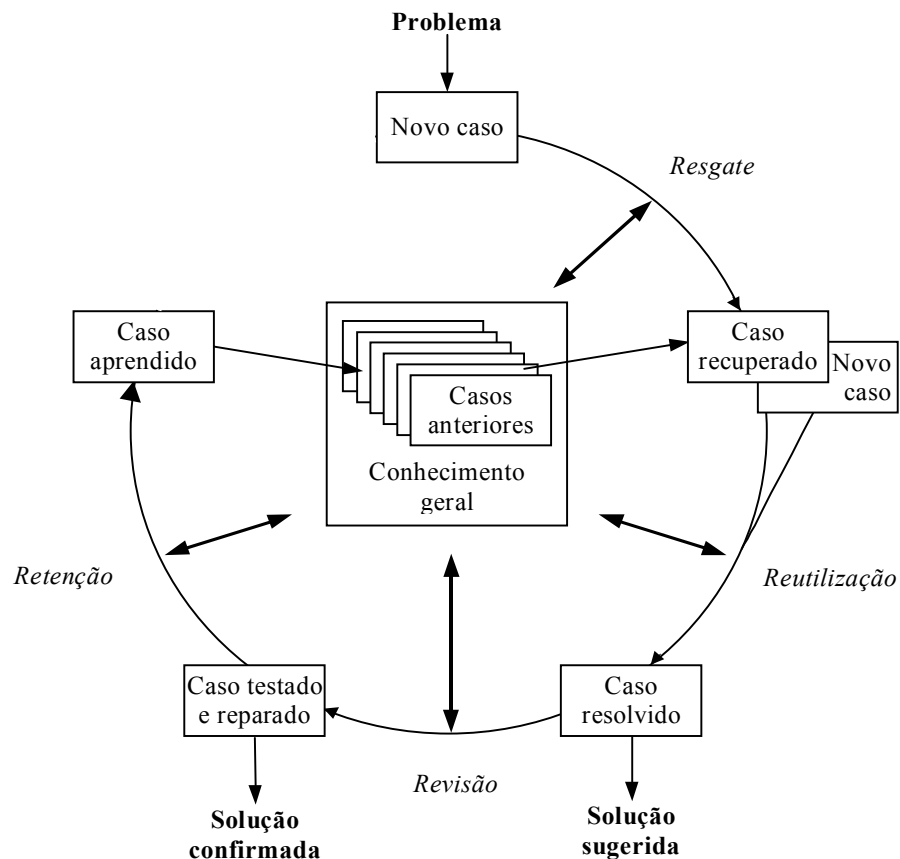


Figura 1 - Ciclo RBC (AAMODT E PLAZA, 1994)

O objetivo do RBC é a reutilização de soluções conhecidas no contexto de um problema novo, de solução ainda desconhecida. Em função disso, a determinação de exemplos de casos adequados, que não precisam necessariamente ser idênticos à situação atual, é um dos problemas centrais desta técnica da IA (WANGENHEIM E WANGENHEIM, 2003).

### 2.3. Cálculo de similaridade

Sendo a determinação de casos adequados um dos problemas centrais em RBC, a eficácia de enfoques baseados em casos depende, essencialmente, da escolha de um conceito de similaridade adequado ao domínio de aplicação e à estrutura dos casos usados. Este conceito de similaridade deve permitir a estimativa da utilidade de um caso com base na similaridade observada entre a descrição do problema atual e a contida no caso.

A medida de similaridade é a formalização de uma determinada filosofia de julgamento de semelhança através de um modelo matemático concreto. Existem várias técnicas matemáticas utilizadas para o cálculo da similaridade entre casos. Qualquer uma delas, no entanto, costuma exibir as seguintes propriedades (MARTINS, 2000):

- Reflexividade: um caso é sempre similar a si mesmo;
- Simetria: se o caso  $A$  é similar ao caso  $B$ , então o caso  $B$  é sempre similar ao caso  $A$ ;
- Não transitividade: se o caso  $A$  é similar ao caso  $B$  e  $B$  é similar ao caso  $C$ , nem sempre se pode garantir que o caso  $A$  é similar ao caso  $C$ .

Existem duas classes de similaridade quanto à sua abrangência: a similaridade global e a similaridade local. Tipicamente, a similaridade global entre o caso alvo ( $Q$ ) e um caso fonte ( $C$ ), ambos descritos por  $n$  atributos, e denotada por  $SIM(Q,C)$ , está baseada na computação das similaridades locais (denotada por  $sim_n(q_n, c_n)$ ) entre cada atributo destes dois casos e pode ser expressa pela equação (1) abaixo (MARTINS, 2000):

$$SIM(Q,C) = f(sim_1(q_1, c_1), sim_2(q_2, c_2), \dots, sim_n(q_n, c_n)) \quad (1)$$

Uma classe particular de métrica para computar essa similaridade pode ser obtida pela soma ponderada das similaridades locais para cada atributo conforme a expressão que segue:

$$SIM(Q,C) = \frac{\sum_{i=1}^n W_i \times sim(q_i, c_i)}{\sum_{i=1}^n W_i} \quad (2)$$

onde  $i$  designa qualquer atributo particular de  $1$  até  $n$ ;  $W$  significa um peso medidor da importância do atributo  $i$ ;  $n$  significa a quantidade total de atributos em cada caso;  $sim(q_i, c_i)$  designa a similaridade particular ou local entre o valor do  $i$ -ésimo atributo do caso alvo ( $Q$ ) e do caso ( $C$ ) na base de casos.

A identificação das características relevantes de um caso vai determinar a sua similaridade com outros casos. Essas características são definidas como índices na recuperação de casos. Num sistema de RBC, os índices representam uma interpretação de uma situação específica e indicam os casos na memória que são mais similares a um caso de entrada usado para pesquisa (BRAGA JR., 2001).

A escolha correta dos índices, em qualquer sistema, depende do que é considerado significativo para o sistema (SCHANK, 1982). Os índices precisam ser cuidadosamente selecionados para que apenas os casos mais úteis possam ser recuperados de forma rápida, eficiente e precisa.

A similaridade é o ponto crucial do RBC, pois é a partir dela que todo o processo de raciocínio se fundamenta, tornando esta técnica viável (KOLODNER, 1993). A similaridade do caso a ser solucionado (caso alvo) é avaliada com relação aos casos fontes. O que faz um caso ser ou não similar é a semelhança das características ou atributos que realmente representam o conteúdo e o contexto da experiência em questão (THÉ, 2001).

## 2.4. Aprendizado de máquina

O campo de Aprendizado de Máquina (AM) preocupa-se com a questão de como construir programas que automaticamente melhoram seu desempenho com a experiência. Diversas aplicações de AM foram desenvolvidas com sucesso, como programas de mineração

de dados que aprendem a detectar transações fraudulentas de cartão de crédito; sistemas de filtro de informações que aprendem as preferências de leitura dos usuários; veículos autônomos que aprendem como dirigir em vias públicas; sistemas de diagnóstico médico, entre outros. Ao mesmo tempo, têm acontecido importantes avanços na teoria e nos algoritmos que foram as bases deste campo (MITCHEL, 1997; GOLBERG, 1989).

Dentre as diversas técnicas de AM podemos listar Aprendizado por Árvores de Decisão, Redes Neurais Artificiais, Redes Bayesianas e Algoritmos Genéticos.

#### **2.4.1. Aprendizado por árvores de decisão**

Um dos mais práticos métodos de aprendizado indutivo é o aprendizado por árvore de decisão, que é um método para classificar instâncias, ordenando-as da informação base (raiz), passando por nós (bifurcações representadas pelo teste de um atributo), ramificando (galhos) até chegar ao último nível (folha) (MITCHEL, 1997).

O algoritmo básico de uma árvore de decisão avalia estatisticamente cada atributo para determinar como ele, por si só, classifica os dados de treinamento, selecionando o melhor atributo para ser utilizado no primeiro nó (raiz). A partir daí, um ramo descendente é criado com os valores possíveis para este atributo e novamente é aplicado o teste com os dados de treinamento, selecionando mais uma vez o atributo de melhor resultado. E o processo se repete para cada nó da árvore.

Diversos algoritmos para Árvores de Decisão foram propostos como o CART (BREIMAN, 1984 *apud* LAROSE, 2005), o ID3 (QUINLAN, 1988 *apud* MITCHEL, 1997) e C4.5 (QUINLAN, 1993 *apud* MITCHEL, 1997). Contudo, uma das limitações dos algoritmos de Árvores de Decisão é que trabalham apenas com valores discretos, e, inclusive, para um melhor entendimento humano, podem ser representadas por conjuntos de regras do tipo “se  $x$  então  $y$ ”. Exigem, ainda, um conjunto de dados para treinamento extenso e variado (MITCHEL, 1997; LAROSE, 2005).



### **2.4.2. Redes neurais artificiais**

A inspiração para redes neurais artificiais foi o reconhecimento que sistemas complexos de aprendizado em cérebros animais consistiam de conjuntos de neurônios interligados. Apesar de um único neurônio em particular ser relativamente simples em estrutura, densas redes de neurônios interconectados podiam executar tarefas complexas de aprendizado como classificação e reconhecimento de padrões. Redes Neurais Artificiais (RNA) representam uma tentativa de imitar o tipo não-linear de aprendizado que ocorre nas redes de neurônios encontradas na natureza (LAROSE, 2005).

Mitchel (1997) considera o aprendizado em RNA como apropriado para problemas em que os dados de treinamento correspondem a dados complexos e com ruídos (distorções e informações não consistentes) como informações vindas de câmeras e microfones, ou ainda para representações simbólicas.

No entanto, ao contrário das Árvores de Decisão, que produzem regras intuitivas inteligíveis por não-especialistas, as RNA são relativamente “caixas-pretas” para a interpretação humana, além de requerer um tempo razoável para o treinamento, podendo chegar a várias horas (LAROSE, 2005; AZEVEDO, BRASIL E OLIVEIRA, 2000).

### **2.4.3. Redes bayesianas**

O raciocínio Bayesiano fornece uma abordagem probabilística para a inferência, sendo baseado na premissa de que as quantidades de interesses são governadas pelas distribuições de probabilidade, e que decisões ótimas podem ser feitas através da análise sobre estas probabilidades junto com os dados observados (MITCHEL, 1997).

As principais características do método Bayesiano de aprendizado são: cada exemplo de treinamento observado pode aumentar ou diminuir a probabilidade estimada de que uma hipótese é correta; o conhecimento anterior pode ser combinado com os dados observados para determinar a probabilidade final de uma hipótese; podem acomodar hipóteses com

predições probabilísticas (não precisas); novas instâncias podem ser classificadas pela combinação de predições de múltiplas hipóteses, ponderadas pelas suas probabilidades; e mesmo em casos em que os métodos Bayesianos não são considerados ideais, eles podem fornecer um padrão de decisão ótima, servindo como medida para os métodos práticos.

Contudo, segundo Mitchel (1997), uma dificuldade prática da aplicação de métodos Bayesianos é que eles tipicamente requerem conhecimento inicial de algumas probabilidades. Outra dificuldade prática é o significativo custo computacional requerido para determinar a hipótese ótima no caso geral, ainda que, em certas situações especiais, este custo computacional possa ser significativamente reduzido.

#### **2.4.4. Aprendizado de máquina baseado em genética**

A teoria para o Aprendizado de Máquina Baseado em Genética (AMBG) foi fundamentada por Holland (1962 *apud* GOLDBERG, 1989). Com esta base teórica e o papel fundamental da recombinação, mais sugestões concretas apareceram para a criação de uma seqüência de esquemas de processadores, até a criação da primeira aplicação de um sistema classificador, ou Sistema Cognitivo (SC), por Holland e Reitman em 1978, que foi chamado de *Cognitive System Level One* ou, simplesmente, CS1 (GOLDBERG, 1989).

Um sistema classificador é um sistema de AM que aprende sintaticamente regras simples em forma de *strings*, chamadas de classificadores, para guiar sua performance em um ambiente arbitrário (GOLDBERG, 1989). Desde o primeiro sistema classificador, diversos pesquisadores expandiram e aplicaram estas idéias nas mais variadas maneiras, conforme se pode ver na Tabela 1.

Um sistema classificador é composto de (i) sistema de mensagens e regras; (ii) sistema de distribuição de crédito; e (iii) Algoritmo Genético (GOLDBERG, 1989).

Tabela 1 - Aplicações de AM Baseadas em Genética (GOLDBERG, 1989)

Ano	Pesquisadores	Descrição
<b>BIOLOGIA E MEDICINA</b>		
1984	Rada, Rhine e Smailwood	Tentativa de um sistema AMBG para diagnóstico médico.
1987	Bickel e Bickel	Desenvolvimento de um sistema de AMBG para o diagnóstico médico.
1987	Wilson	Proposta para simulações de morfogênese usando um sistema classificador.
<b>NEGÓCIOS</b>		
1986	Frey	Classificação de arquitetura usando um sistema cognitivo.
1986	Thompson e Thompson	Algoritmo Genético usado para buscar por conjuntos de regras para prever lucratividade de empresas.
1987	Greene e Smith	Sistema AMBG aprende regras que descrevem preferências do consumidor.
<b>CIÊNCIA DA COMPUTAÇÃO</b>		
...		
1985	Cramer	Algoritmo Genético aprende tarefas de multiplicação usando conjunto de instruções similar ao <i>assembler</i> .
...		
<b>ENGENHARIA E OPERAÇÃO</b>		
...		
1985	Kuchinski	Algoritmo Genético busca regras de sistema de gerenciamento de batalhas.
...		
<b>TÉCNICAS HÍBRIDAS</b>		
1987	Oosthuizen	Sistema híbrido de AM integra conexionismo, indução gráfica e Algoritmo Genético.
<b>APRENDIZADO DE MÁQUINA</b>		
...		
1986	Holland, Holyoak, Nesbitt e Thagard	Publicação do livro "Indução".
...		
<b>IMPLEMENTAÇÕES PARALELAS</b>		
1987	Robertson	Implementação de um sistema classificador em uma Máquina de Conexão.
<b>CIÊNCIAS SOCIAIS</b>		
1978	Holland e Reitman	Primeiro sistema classificador (CS1) aprende duas tarefas cíclicas.
...		
1987	Fujiko e Dickinson	Algoritmo Genético aprende código LISP para resolver o problema "dilema dos prisioneiros".

Um sistema de mensagens e regras de um sistema classificador é um tipo especial de sistema de produção, que usa regras como seu único algoritmo, geralmente na forma: se <condição> então <ação>. A distribuição de crédito ocorre quando muitos sistemas classificadores tentam atribuir pontos, taxas ou notas para classificadores individuais de acordo com o papel do classificador em conseguir recompensa, ou retorno positivo, do

ambiente. Contudo, deve ainda haver uma maneira de se incluir novas e, possivelmente, melhores regras no sistema. É neste ponto que entra o terceiro componente, o Algoritmo Genético, o qual será abordado em detalhe na seção seguinte.

## **2.5. Algoritmos genéticos**

### **2.5.1. Computação evolucionária**

Em 1858, Charles Darwin e Alfred Russel Wallace lançaram as bases da Teoria da Evolução, causando uma revolução tanto do pensamento biológico quanto da filosofia humana (FOGEL, 1995). Desde então, esta teoria é uma das mais aceitas pelo mundo científico, sustentando que o processo de evolução natural das espécies se dá pela seleção dos indivíduos mais aptos ao ambiente em que vivem, e assim têm maiores chances de sobreviver e gerar filhos, que herdam as características dos pais.

Computação evolucionária é o nome genérico dado a métodos computacionais inspirados na teoria da evolução e os algoritmos usados em computação evolucionária chamam-se Algoritmos Evolucionários (AE) (BARRETO, 1996).

Atualmente, os AE mais conhecidos são: Algoritmos Genéticos, Programação Genética e Estratégias Evolucionárias.

### **2.5.2. Algoritmos genéticos**

Os Algoritmos Genéticos (AG) foram desenvolvidos inicialmente por Holland (1975), no trabalho intitulado *Adaption in Natural and Artificial Systems*. De acordo com Goldberg (1989) os AG constituem um método de busca adaptativa em que se busca a solução ótima para um dado problema. São baseados na analogia entre otimização e evolução natural das espécies, combinando os conceitos de adaptação seletiva e sobrevivência dos indivíduos mais capazes.

A evolução, segundo o Neo-Darwinismo (FOGEL, 1995), se dá com a aplicação dos processos de reprodução, mutação, competição e seleção interagindo nas populações, geração após geração. Do mesmo modo, AG utilizam operadores que tentam simular estes processos. Segundo Grefenstette (1986), AG é um procedimento iterativo que mantém uma população de estruturas (chamadas indivíduos ou *string*), que representam possíveis soluções de um determinado problema. A cada incremento temporal (chamado geração), os indivíduos na população atual são avaliados de acordo com o valor de seu *fitness* (desempenho ou aptidão) para solução do problema. Tendo como base essa avaliação, uma nova população de soluções candidatas é formada utilizando Operadores Genéticos (OG) específicos, tais como *crossover* e mutação.

### **2.5.3. Problema de otimização**

O problema endereçado por AG é pesquisar um conjunto de hipóteses candidatas para identificar a melhor hipótese. Em AG, a “melhor hipótese” é definida como aquela que otimiza uma medida numérica pré-definida para o problema em mãos, chamada *fitness*. (MITCHEL, 1997).

Os AG podem ser enquadrados em grande parte dos problemas científicos a serem formulados como problemas de busca e otimização.

Basicamente, existe uma série de fatores influenciando o desempenho de um dado sistema. Tais fatores podem assumir um número limitado ou ilimitado de valores, podendo ser sujeitos a certas restrições. O objetivo é encontrar a melhor combinação dos fatores, ou seja, a combinação de fatores que proporcione o melhor desempenho possível para o sistema em questão. Em termos técnicos, o conjunto de todas as combinações possíveis para os fatores constitui o chamado espaço de busca. Não é difícil perceber que existe uma dualidade entre os conceitos de busca e otimização, de tal modo que todo problema de busca pode ser

considerado um problema de otimização e vice-versa (TANOMARU, 1995; GOLDBERG, 1989).

Os AG fazem parte da classe dos métodos probabilísticos de busca e otimização, apesar de não serem aleatórios. Os AG usam o conceito de probabilidade, mas não são simples buscas aleatórias. Ao contrário, AG tentam direcionar a busca para regiões onde é provável que os pontos ótimos estejam (AZEVEDO, 1999).

Além disso, em relação às técnicas de busca convencionais, os AG diferem nos seguintes pontos (AZEVEDO, 1999):

- A busca da melhor solução para o problema é feita sobre uma população de pontos, e não sobre um único ponto, reduzindo sensivelmente o risco da busca ficar estagnada sobre um máximo (ou mínimo) local;
- AG realizam uma busca cega. A única exigência é o conhecimento do valor da função de custo (ou objetivo) de cada indivíduo. Não há necessidade de qualquer outra informação ou heurística dependente do problema;
- AG usam operadores estocásticos, e não regras determinísticas, para guiar uma busca altamente exploratória e estruturada, onde informações acumuladas nas iterações (gerações) anteriores são usadas para direcionar essa busca.

Apesar de sua simplicidade, os resultados obtidos com a aplicação do método, segundo Goldberg (1989), permitem concluir que AG são um método de busca robusto, eficiente e eficaz em uma grande variedade de problemas.

#### **2.5.4. Terminologia**

Em nível biológico, um indivíduo é formado por um conjunto de cromossomos. No entanto, pode-se fazer uma analogia entre indivíduo e cromossomo, tendo em vista que um indivíduo pode ser formado por apenas um cromossomo, o que é comum em AG. Por isso, os dois termos são utilizados indistintamente (AZEVEDO, 1999).

Assim, um indivíduo é definido por uma *string* usando um alfabeto finito, de modo que cada *string* represente um conjunto de valores para os parâmetros do problema. Um exemplo de alfabeto é o conjunto  $\{0,1\}$ , ou o conjunto de números inteiros. Deste modo, cada posição da *string* representa um gene.

O cromossomo é composto de genes, sendo que cada gene possui um local fixo no cromossomo, este local é denominado de *locus*. Cada gene pode assumir certo valor pertencente a certo conjunto de valores, os quais são denominados de alelo. Em termos de AG, o gene é cada um dos caracteres da *string* (no caso binário, pode ser denominado *bit*) e o *locus* é de posição do caractere (ou *bit*, no caso binário) no indivíduo ou *string*. Já o termo alelo, refere-se ao conjunto de valores possíveis de serem atribuídos a um determinado caractere ou posição da *string*, ou seja, por exemplo, o conjunto de números binários  $\{0,1\}$  ou inteiros  $\{0,1,2,3,4,5,6,7,8,9\}$  (AZEVEDO, 1999).

O conjunto de cromossomos, genes e alelos denomina-se genótipo e as características conferidas por este, denota-se de fenótipo. Em termos de AG, o genótipo é a variável independente  $x$  e o fenótipo a variável dependente ou função  $f(x)$ . Uma comparação entre a terminologia utilizada em AG e a terminologia utilizada na biologia encontra-se na Tabela 2.

**Tabela 2 - Terminologia dos AG x Biologia (LOPES, 1996; CORRÊA, 2000)**

BIOLOGIA	ALGORITMOS GENÉTICOS
cromossomo	indivíduo, ou <i>string</i>
gene	gene, caractere, ou <i>bit</i> (no caso binário)
alelo	valor do gene, ou do <i>bit</i> (no caso binário)
<i>locus</i>	posição de um gene específico no indivíduo ou <i>string</i>
genótipo	estrutura, indivíduo candidato
fenótipo	solução, valor da função $f(x)$ para um dado indivíduo

Segundo Lacerda e Carvalho (1999), embora a cadeia de *bits* seja a forma mais tradicional, nem sempre ela é a melhor. Um cromossomo é uma estrutura de dados e, dentre outras opções também pode ser um vetor.

Normalmente, os AG trabalham com um conjunto de indivíduos (população), no qual cada elemento é candidato a ser a solução desejada. A função a ser otimizada é o ambiente no qual a população inicial vai ser posta. Espera-se, que através dos mecanismos de evolução das espécies e da genética, somente os mais aptos se reproduzam e também que cada nova geração esteja mais apta ao ambiente (função a ser otimizada).

O grau de *fitness* (aptidão) de cada indivíduo é obtido pela avaliação de tal indivíduo através da função a ser otimizada. Se o objetivo for maximizar, o *fitness* é diretamente proporcional ao valor da função (TANOMARU, 1995). Caso o objetivo seja a minimização da função, o *fitness* será inversamente proporcional ao valor da função. Contudo, o termo minimização não é bem aceito por alguns pesquisadores por não ter inspiração biológica, haja vista que somente o mais apto é que deve sobreviver (ALVES, 2002; TSUNODA, 2004).

Assim, quando já se tem realizado o teste de todos os indivíduos da população na função a ser otimizada, obtém-se o *fitness* para cada um, ou seja, o seu grau de aptidão.

A próxima geração será uma evolução da anterior e para que isso ocorra os mais aptos (melhor *fitness*) deverão possuir maior probabilidade de serem selecionados para dar origem a nova geração. Com isso, se o processo for bem conduzido, espera-se que a nova geração seja, em média, melhor do que a que lhe deu origem.

A seleção dos indivíduos da geração anterior que vão participar da formação da nova geração pode ser realizada através da roleta ou *ranking* (CORRÊA, 2000; ALVES, 2002; SILVA, 2005b). Nos dois métodos, os indivíduos são ordenados de acordo com sua nota (*fitness*), possuindo mais chances de serem selecionados, aqueles com as maiores notas. A diferença é que a roleta pondera utiliza como faixa de seleção a posição acumulada, e não a nota acumulada, aumentando ainda mais as chances dos indivíduos com notas maiores.

Também pode ser feita a Seleção por Torneio. Neste caso são escolhidos aleatoriamente (com probabilidades iguais)  $n$  cromossomos da população e o cromossomo com maior aptidão



é selecionado para a população intermediária. O processo repete-se até preencher a população intermediária. De acordo com Lacerda e Carvalho (1999), utiliza-se geralmente o valor de  $n=3$ .

A escolha dos melhores indivíduos para participar da reprodução faz com que a média da população caminhe na direção mais promissora da solução desejada (MACHADO, 1991; TANOMARU, 1995).

Realizada a seleção, o próximo passo é a aplicação dos mecanismos de busca, como também são conhecidos os já citados OG. Entre tais mecanismos, os mais comumente empregados em AG são *crossover* e mutação. Estes operadores serão descritos com maiores detalhes na próxima seção.

Um outro ponto a ser mencionado é com respeito aos parâmetros do AG, ou seja, os valores que influenciam o desempenho do AG. Seguindo a relação proposta por Austin (1990 *apud* AZEVEDO, 1999) estes parâmetros são: tamanho da população, taxa de operadores, intervalo de geração, estratégia de seleção e fator de escalonamento.

O tamanho da população de cromossomos afeta o desempenho global dos AG. Uma população pequena é insuficiente para cobrir o espaço de busca do problema. Uma população grande é mais representativa do domínio, além de evitar a convergência prematura para soluções locais, em vez de soluções globais.

As taxas de operadores medem a frequência com que cada tipo de OG é utilizado. Representam também a influência que cada tipo de OG exerce sobre a população de cromossomos. Por exemplo, se a taxa do operador de *crossover* for muito alta, alguns cromossomos de bom desempenho podem ser removidos mais rapidamente do que a seleção possa desenvolvê-los. Se a taxa de *crossover* for muito baixa, a busca pode estagnar.

O intervalo de geração controla o percentual da população a ser substituído durante cada ciclo de geração. Por exemplo:  $N \times G$  cromossomos da população  $P(t)$  são escolhidos

para serem substituídos na população  $P(t+1)$ . Se o valor de  $G$  for igual a  $I$ , significa que toda a população é substituída durante cada geração.

As estratégias de seleção correspondem aos critérios utilizados para a escolha de cromossomos durante a reprodução.

O fator de escalonamento mede a manutenção da diversidade genética da população de cromossomos durante a evolução. Um cromossomo ou um grupo de cromossomos pode ter uma aptidão bastante forte, a ponto de dominar o processo de reprodução reduzindo a diversidade da população. Uma maneira de se controlar este processo é ordenando os cromossomos, escalonando o seu desempenho para refletir sua aptidão relativa dentro da população, e utilizando as operações genéticas de mutação para reduzir a homogeneidade da população de cromossomos.

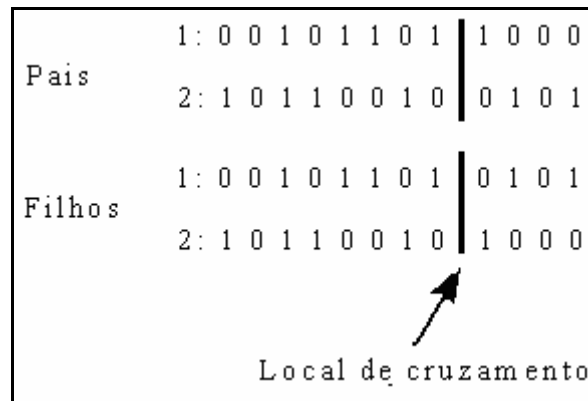
Por último, após a seleção e a aplicação dos OG, tem-se uma nova geração a qual deve ser avaliada, visando comparar o seu grau de *fitness* em relação à geração anterior. Caso tal geração não esteja apta o suficiente, deve-se repetir o processo de seleção e reprodução até que o grau de *fitness* seja aceitável (ALGARVE, 1994).

### **2.5.5. Operadores genéticos**

Os dois principais operadores conhecidos são o *crossover* e a mutação.

O *crossover* é uma operação que simula a reprodução, típica de seres mais evoluídos (ALGARVE, 1994). Ele combina as características de duas estruturas pais (cromossomos dos pais) para formar dois descendentes (filhos) similares. O operador *crossover* opera por selecionar uma localização aleatória na *string* dos pais (ponto de *crossover*) e concatenar o segmento inicial de um pai com o segmento final do segundo pai, de forma a criar um novo filho. Um segundo filho é simultaneamente gerado usando os segmentos restantes dos dois pais, como mostrado na Figura 2. Este é o operador mais simples e é chamado *crossover* de um ponto (*one-point crossover*). Há, ainda, muitas outras técnicas de *crossover*, como é o

caso do *crossover* de dois pontos (*two-point crossover*) e o *crossover* uniforme (GARIS, 1992; SRINIVAS E PATNAIK, 1994). Contudo, não há consenso sobre qual é a melhor técnica a ser usada para corresponder aos objetivos esperados (AZEVEDO, 1999).



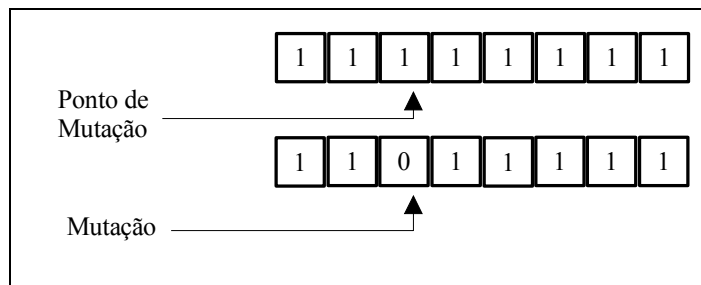
**Figura 2 - Crossover de um ponto**

Portanto, para a aplicação do *crossover*, geram-se dois números aleatórios: o primeiro entre 0 e 1 indicará a probabilidade de ocorrer *crossover* e o segundo, o local da realização do *crossover*. Caso o primeiro número gerado seja inferior ao definido pelo usuário como probabilidade de *crossover*, realiza-se o *crossover* propriamente dito, e caso contrário, copiam-se os pais para a nova geração. Segundo Tanomaru (1995) este índice é geralmente de 60%, posição similar à de Lacerda e Carvalho (1999) que indicam a taxa de *crossover* variando entre 60% e 90%. O segundo número aleatório, gerado apenas no caso da ocorrência de *crossover*, indicará a posição de corte do cromossomo para efetuar o *crossover*. Para tanto, o número aleatório gerado deverá estar entre 1 e  $g-1$ , onde  $g$  é o número de genes ou *bits* do cromossomo.

A mutação consiste em uma pequena alteração em determinado ponto da *string* da cadeia de cromossomos, gerando um indivíduo diferente (SILVA, 2005b).

O parâmetro de probabilidade de mutação determina a ocorrência desta perturbação, e esta alteração ocorre de forma que cada gene em cada cromossomo é um candidato à mutação. O ponto da mutação é escolhido de maneira aleatória.

Esta probabilidade é mantida usualmente em um valor baixo para evitar a perda de um número grande de cromossomos bons (PARK, 1994). O operador de mutação pode ser implementado de várias maneiras. A codificação binária de *string* é o modo mais fácil para executá-la.



**Figura 3 - Mutação em codificação binária**

A tarefa da mutação em AG tem sido a de restituir a perda ou material genético inexplorado na população, com o objetivo de prevenir a convergência prematura do AG para soluções “sub-ótimas” (SRINIVAS E PATNAIK, 1994)

Dentre os principais mecanismos de alteração genética que recebem a denominação global de mutação, destacam-se: troca simples, translocação, inversão, deleção e adição.

Na adição, ocorre a inserção de mais um gene na cadeia e na deleção é justamente o oposto, ou seja, ocorre a retirada de um gene da cadeia. Geralmente estes mecanismos não são utilizados em algoritmos genéticos, pois alteram o comprimento da cadeia do cromossomo (AZEVEDO, 1999).

A troca simples consiste de um erro de cópia de um ou mais genes da cadeia. Se um gene for considerado como sendo um *bit* com valor lógico 1, a ocorrência de troca simples levaria este *bit* (gene) para nível lógico 0 e se fosse 0 levaria para 1. Já a inversão consiste na retirada e inserção de um pedaço da cadeia porém, na ordem inversa da que foi retirada. Por sua vez, a translocação é a retirada uma parte do cromossomo e a sua recolocação em outra posição do mesmo cromossomo. Estes três últimos mecanismos não alteram o comprimento

original da cadeia. No entanto, como na maioria dos trabalhos com AG, este também usa o termo mutação como sinônimo de troca simples (HOLLAND, 1975; ALGARVE, 1994).

A probabilidade de ocorrer mutação é sempre bem menor que a de ocorrer *crossover*. Segundo Srinivas e Patnaik (1994), existe um compromisso entre os três parâmetros de controle do AG simples: tamanho da população, probabilidade de *crossover* e probabilidade de mutação. Muitos autores têm proposto valores para estes parâmetros, visando garantir uma boa performance do AG, porém estes valores ainda fazem parte de uma heurística. Lacerda e Carvalho (1999) sugerem uma taxa de mutação entre 0,1% e 5%, pequena porque cada mutação destrói a informação do cromossomo, mas o suficiente para assegurar a diversidade dos “filhos” para a população gerada.

#### **2.5.6. Utilização de números reais**

Como apresentado anteriormente, o cromossomo (indivíduo, *string*) é representado por um vetor de dados. A representação binária é utilizada desde os trabalhos pioneiros de Holland (1975). Contudo, alguns pesquisadores (MICHALEWICZ, 1992; HAUPT E HAUPT, 1998) têm feito experimentos favoráveis à representação com números reais.

As vantagens alegadas por estes pesquisadores são que a representação do cromossomo com números reais é compreendida mais naturalmente pelo ser humano do que aquela usando uma cadeia de *bits*. Além disso, a representação por valores reais requer menos memória (o cromossomo fica bem menor). Os defensores desta representação também alegam que além de cromossomo longos na memória, os operadores não são uniformes. Por exemplo, mutação nos primeiros *bits* do gene afeta mais a aptidão do que mutação nos últimos *bits* do gene (LACERDA E CARVALHO, 1999).

Outra vantagem alegada para a representação real é a possibilidade de criar novos operadores e variações dos operadores existentes. Os operadores convencionais (*crossover* de *n*-pontos e uniforme) funcionam bem na representação binária, mas na representação real eles

basicamente trocam valores dos parâmetros e, portanto, não criam informações novas. Neste caso, melhor então seria usar operadores aritméticos como o *crossover* média (DAVIS, 1991), no qual é produzido apenas um filho, cujos genes são obtidos pela média entre os genes da mesma posição dos cromossomos pais.

A fórmula para cálculo dos alelos no *crossover* média é apresentada pela equação a seguir, onde  $a$  é o alelo do filho,  $i$  é o número do gene,  $m$  é o alelo do indivíduo mãe e  $p$  o alelo do indivíduo pai:

$$a_i = \frac{(m_i + p_i)}{2} \quad (3)$$

Uma variação do *crossover* média é o *crossover* média geométrica, que também gera apenas um filho. Neste caso, o valor de cada gene do cromossomo é calculado por:

$$a_i = \sqrt{m_i + p_i} \quad (4)$$

Existem outros operadores como o *blend crossover* ou *crossover* BLX- $\alpha$  (ESHELMAN E SCHAFFER, 1993 *apud* LACERDA E CARVALHO, 1999) e os *crossovers* com aplicação de um número aleatório escolhido de uma distribuição uniforme no intervalo entre 0 e 1 (RADCLIFFE, 1991).

E assim como para os operadores de *crossover*, existem diversas variações para os operadores de mutação, como a mutação randômica, que é uma simples substituição de um gene por um número escolhido aleatoriamente no intervalo permitido para o problema, e a mutação *creep*, que adiciona ao gene um pequeno número aleatório (LACERDA E CARVALHO, 1999).

## 2.6. Projeto IACVIRTUAL

O Projeto IACVIRTUAL (ALMEIDA et al., 2003a, 2003b, 2004a, 2004b) contempla a modelagem e a implementação de um consultório médico virtual via WEB utilizando técnicas de IA e foi concebido para atender pacientes interessados em acompanhar seu histórico médico, especialistas interessados num ambiente de apoio à decisão em diagnóstico e tratamento de seus pacientes, bem como estudantes interessados na aprendizagem por meio do acompanhamento dos casos médicos disponíveis. Este projeto está em desenvolvimento no Mestrado em Gestão do Conhecimento e da Tecnologia da Informação (MGCTI), na área de Tecnologia da Informação em Saúde, e está sendo piloto para duas das áreas de excelência do Centro de Ciência e Tecnologia (CCT): TI na saúde, oriunda do MGCTI, e Engenharia Biomédica, como curso de graduação na UCB. Além disso, este projeto é considerado no Programa de Pós-Graduação na UCB como estruturante, devido à sua característica de englobar muitas tecnologias de ponta, várias áreas de aplicação na saúde, pessoal multidisciplinar e alocação de estudantes de graduação e pós-graduação. Ele conta com o apoio de várias instituições brasileiras e internacionais, clínicas, empresas de equipamentos médicos, especialistas médicos, pesquisadores, professores, estudantes de Iniciação Científica, e tem sido objeto de projetos de conclusão de cursos de graduação e de pesquisas de programas de pós-graduação das instituições participantes.

Como apresentado na Figura 4, o IACVIRTUAL é composto pelos seguintes módulos:

- O Módulo de Dados oferece uma base estruturada de dados e de casos, incluindo os dados clínicos, os elementos para o processamento de imagens e de sinais e para telemedicina.
- O Módulo de Interface, além dos elementos clássicos de hipermídia, oferece recursos de realidade virtual.

- O Módulo de Usuário prevê três níveis/tipos de usuários: (i) o especialista; (ii) o paciente e (iii) o estudante. A interface da ferramenta se adapta de acordo com o perfil do usuário.

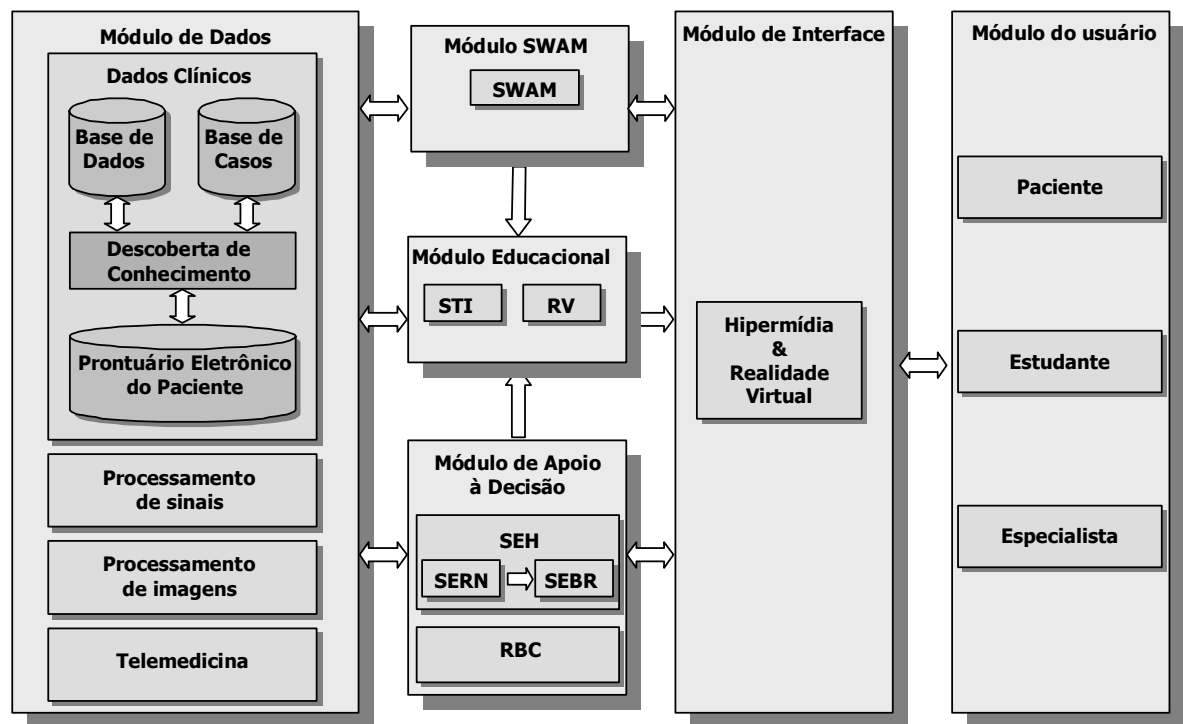


Figura 4 - Projeto IACVIRTUAL

- O Módulo SWAM, Sistema Web para Análise de Mamografia Digital, tem o objetivo de apoiar no diagnóstico médico na identificação da existência de um tumor maligno e seu tratamento.
- O Módulo Educacional, compreende dois sub-módulos: Sistema Tutor Inteligente (STI) e Realidade Virtual (RV). Ambos serão conectados devido à metodologia empregada pelo STI que prevê um processo para organização do conteúdo a ser ensinado de forma que uma das atividades a ser proposta considera a possibilidade de inclusão de simulações de ambientes de saúde, corpo do paciente, órgãos do paciente, cirurgias, etc.



- O STI, concebido segundo a arquitetura MATHEMA (COSTA, LOPES E FERNEDA, 1995), busca promover interações cooperativas entre um aprendiz e uma sociedade de agentes artificiais do tutor com o objetivo de promover a aprendizagem do estudante através do Sistema Tutor Inteligente.
- O RV, nesta etapa, consiste na implementação de um conjunto de funcionalidades que permitirão importar modelos volumétricos, codificados nos principais padrões existentes, e visualizar esses modelos através de uma interface gráfica interativa, onde o usuário disporá de ferramentas básicas de navegação tridimensional.
- O Módulo de Apoio à Decisão dá suporte ao diagnóstico médico e, indiretamente, subsidia o Módulo Educacional quanto à resolução de problemas. Ele é composto pelos sub-módulos Sistema Especialista Híbrido (SEH) e Raciocínio Baseado em Casos (RBC):
- No SEH o conhecimento é representado por meio de um formalismo que integra regras de produção, redes neurais artificiais, algoritmo genético e lógica *fuzzy* (ROJAS, 2003).
- No RBC o conhecimento é representado por uma base de casos (SILVA et al., 2004).

Os resultados desta pesquisa serão aplicados neste sub-módulo RBC, o qual terá incorporada a funcionalidade de suporte à definição de pesos de características dos casos por AG.

## **3. Metodologia**

---

### **3.1. Classificação da pesquisa**

Do ponto de vista da abordagem do problema, esta é uma pesquisa quantitativa, porque seus resultados permitirão a classificação e a análise científica, haja vista que será realizada sobre uma base de casos de Cardiologia com diagnósticos reais de Cardiopatia Isquêmica.

Quanto aos meios, por se tratar de simulações feitas em computador, classifica-se como pesquisa de laboratório.

Do ponto de vista da natureza e quanto aos fins, a presente pesquisa classifica-se como metodológica e aplicada (VERGARA, 2000; MORESI, 2003). Metodológica porque o estudo a que se refere está associado a caminhos, formas, maneiras, procedimentos para atingir determinado fim. Também é aplicada porque é fundamentalmente motivada pela necessidade de resolver problemas concretos, tendo, portanto, finalidade prática, ao contrário da pesquisa pura, motivada basicamente pela curiosidade intelectual do pesquisador e situada, sobretudo, no nível da especulação.

### **3.2. Suposições e hipóteses**

Esta pesquisa pretende testar a seguinte hipótese:

A melhor solução do AG determinará um conjunto de pesos para a função de similaridade global do RBC cuja acuidade auxiliará o especialista médico na definição final do conjunto de pesos a ser utilizado no módulo RBC.

### **3.3. Coleta e análise de dados**

Como base para os experimentos, será utilizada a atual base de casos do módulo RBC do Projeto IACVIRTUAL.

Esta base possui 529 casos de cardiopatia isquêmica, cedidos pela Faculdade Baiana de Cardiologia, cujos dados já foram preparados e modelados com a qualidade e informações necessárias para o módulo RBC.

### **3.4. Delimitação do estudo**

O módulo RBC do Projeto IACVIRTUAL será implementado numa interface amigável, permitindo a simulação de um consultório, cujas funções de busca e recuperação serão otimizadas. Contudo, seu modelo não será alterado.

Tratar-se-á somente dos pesos para a função de similaridade global. As funções de similaridade local estão fora do escopo deste estudo.

### **3.5. Resultados esperados**

Esperam-se os seguintes resultados deste trabalho:

- Elaboração do modelo GAC-BRASIL, que definirá um AG para determinação de pesos para funções de similaridade global em RBC.
- Implementação do módulo AG, baseado no modelo GAC-BRASIL, para o cálculo automático de pesos das características dos casos do modelo RBC do Projeto IACVIRTUAL;
- Otimização da ferramenta de busca e recuperação de casos do módulo RBC do Projeto IACVIRTUAL;
- Validação do método proposto numa base de casos reais na área de Cardiologia, contendo diagnósticos de Cardiopatia Isquêmica;
- Comparação, com o auxílio de um especialista médico, dos resultados da melhor solução do AG proposto com os resultados da atual definição de pesos do modelo RBC do Projeto IACVIRTUAL.

### 3.6. O modelo GAC-BRASIL

Como explanado anteriormente, o módulo RBC do Projeto IACVIRTUAL calcula a similaridade dos casos correntes com os casos da base de casos através de uma função ponderada de similaridade global. Para este cálculo são utilizados pesos, um para cada característica (similaridade local), que no modelo atual foram atribuídos por um especialista médico da área em estudo.

Para este trabalho, a idéia é definir um modelo e implementar um AG em que cada indivíduo da população seja uma solução para este problema. Ao final, a melhor solução deste AG será aplicada ao módulo RBC e, com o auxílio do especialista médico, os resultados desta solução serão comparados aos resultados obtidos pela atual definição de pesos.

O modelo foi denominado GAC-BRASIL, por ser um Algoritmo Genético aplicado à Similaridade para Raciocínio Baseado em Casos (*Genetic Algorithm applied to Case-Based Reasoning SIMilarity*)

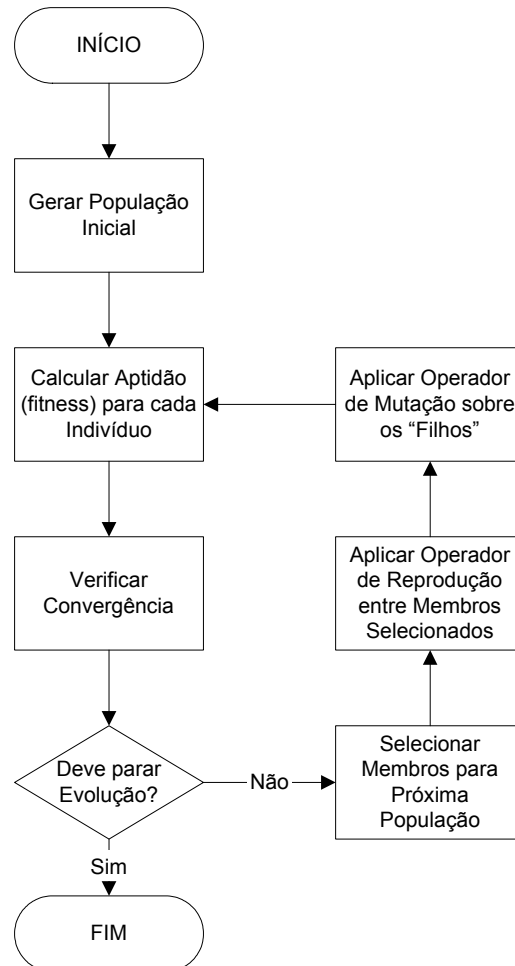
A Figura 5 representa o fluxograma básico do modelo GAC-BRASIL.

#### 3.6.1. Parâmetros e definições

Os parâmetros do AG devem ser totalmente configuráveis para adequação ao problema e para ajustes durante as simulações. Para este trabalho, o GAC-BRASIL utilizará os seguintes parâmetros e definições iniciais:

- Cada indivíduo (cromossomo) é uma proposta de solução para o problema. Assim, deve representar o conjunto de pesos utilizados na função de similaridade global do RBC, com cada gene do cromossomo representando cada um dos pesos.
- A cadeia de cromossomos terá como número de genes a mesma quantidade de pesos utilizados na função de similaridade global do modelo RBC.

- Para a representação dos genes do cromossomo, serão implementadas as opções de utilização de dígitos binários e a representação por números reais.



**Figura 5 - Fluxograma básico do GAC-BRASIL**

- A população inicial será gerada aleatoriamente.
- O número de indivíduos da população também será implementado de forma flexível. De acordo com os experimentos, poder-se-á alterar este parâmetro de modo a obter a melhor relação para a cobertura do espaço de busca.
- O critério de parada do AG também deve ser flexível, permitindo as opções de parar ao atingir um dado número de gerações ou quando não se constatar um percentual mínimo de variação da aptidão média da população. Pode ainda ser utilizada uma

opção mista, onde o AG seria interrompido ao se constatar uma variação mínima da aptidão média, mas somente após um determinado número de gerações.

- Para a seleção, dentre os métodos já explanados no referencial teórico, o *ranking* foi considerado o mais adequado para o problema. Com este método espera-se não atingir prematuramente a convergência e, ao mesmo tempo, permitir atingir a solução otimizada, visto que os membros com maior aptidão terão maior probabilidade de ser selecionados:
  - Os cromossomos serão classificados em ordem decrescente da nota de aptidão, gerando um limite mínimo e máximo para sua escolha;
  - Um número aleatório será escolhido sucessivamente até que o percentual pré-determinado de selecionados seja atingido;
- Serão selecionados dois indivíduos e será aplicado um teste aleatório para verificar a probabilidade de cruzamento. Os pesquisadores estudados e citados no referencial teórico indicam um percentual de 60% a 90%, também parametrizável na aplicação. Caso não sejam indicados para o cruzamento, serão copiados para a população seguinte.
- Caso sejam selecionados para o cruzamento, será utilizado o *crossover* de um ponto para a representação binária, ou o *crossover* média para a representação por números reais. Neste caso, os membros que serão copiados para a próxima população serão os “filhos” (ou o “filho”) resultantes do cruzamento.
- Antes de serem copiados para a próxima população, os filhos serão submetidos a um teste para determinar se sofrerão mutação. Este operador terá a função de garantir a diversidade da população, contudo, deve ser aplicado num baixo percentual da população, para evitar a perda de bons membros. Este percentual é mais um item parametrizável do modelo.

- No caso de representação por números reais, será utilizada a mutação randômica: a troca do valor de um gene por outro escolhido aleatoriamente dentro do intervalo do problema. Na representação por dígitos binários ocorrerá a troca simples: substituição de “0” por “1” ou vice-versa.
- Este processo se repetirá até que seja completada a próxima população, que, como utilizado pela maioria dos autores citados, terá um tamanho fixo.
- Para os novos membros é calculada a aptidão (*fitness*) e, com esta nova geração, o algoritmo retorna ao ponto de verificação do critério de parada.

### 3.6.2. A função *fitness*

Na natureza, a probabilidade de sobrevivência é proporcional à capacidade do indivíduo melhor se adaptar ao meio. Em AG, a definição da nota de aptidão de cada solução é parte fundamental para o sucesso desta técnica, e seu cálculo deve ser adequado ao problema.

Neste trabalho, o melhor indivíduo será considerado a solução para o conjunto de pesos das características da função de similaridade global do RBC. Portanto, a função que calculará a aptidão deve ser uma função que avalie a capacidade do conjunto de pesos de determinar a similaridade de casos, usando a lógica da função de similaridade global atualmente utilizada.

Como base para os experimentos, será utilizada a atual base de casos do módulo RBC do Projeto IACVIRTUAL. Esta base, que compreende 529 casos, foi formada a partir de uma base original de 1.052 casos de cardiopatia isquêmica, cedida pela Faculdade Baiana de Cardiologia, Bahia, Brasil.

Para a função *fitness*, destes 529 casos serão escolhidos dois grupos de casos. O primeiro grupo, com poucos casos, fará o papel de “problemas correntes” do RBC, e o segundo grupo, com um maior número de casos, fará o papel da “base de casos” do RBC.

Assim, a cada geração, para cada um dos indivíduos, será calculada a aptidão (*fitness*) da seguinte maneira:



- A função recebe um indivíduo (que é um conjunto de pesos) como parâmetro de entrada.
- Utilizam-se os genes do indivíduo como os pesos de cada característica local, para calcular a similaridade dos casos do primeiro grupo (problemas correntes) com os casos do segundo grupo (casos da base).

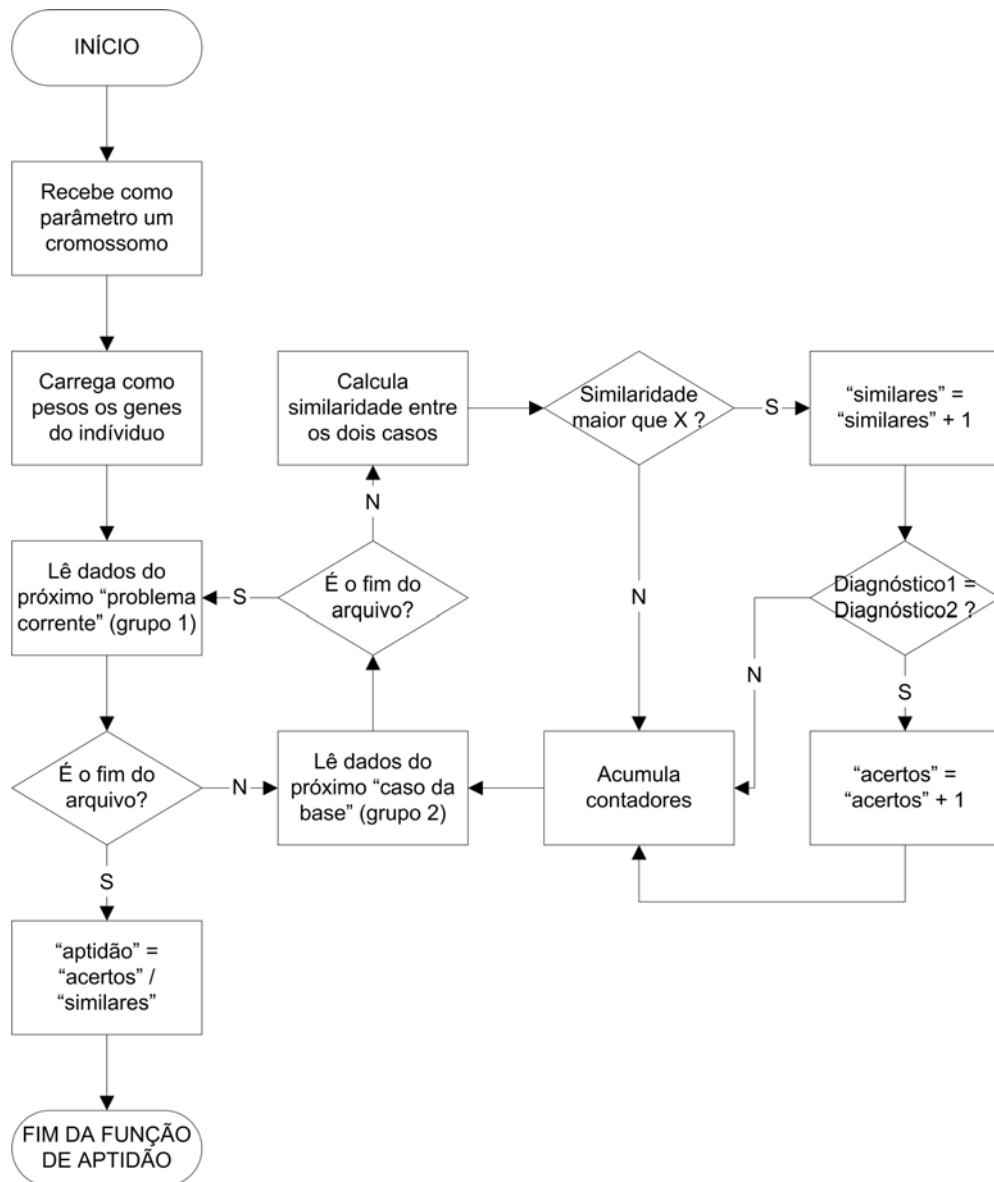


Figura 6 - Fluxograma da função de aptidão (*fitness*)

- Para o resultado de cada uma destas combinações, o caso da base será considerado similar ao problema corrente se a similaridade global atingir um limiar mínimo,

representado por “X” no fluxograma da Figura 6. Este limiar mínimo de similaridade é mais um parâmetro variável para a função do AG.

- Somente para os casos considerados similares, com base no critério acima descrito, será verificado se o diagnóstico real do problema corrente (membro do primeiro grupo) é igual ao diagnóstico real do caso da base (membro do segundo grupo).
- A aptidão do indivíduo (conjunto de pesos recebido como parâmetro) será igual ao percentual dos casos similares cujo com diagnóstico do caso da base seja igual ao do problema corrente.

Esta função avalia se os casos que aquele conjunto de pesos levou a função de similaridade global a considerá-los similares ao problema corrente têm o mesmo diagnóstico do problema corrente.

## **4.Resultados**

---

Um dos objetivos deste estudo é a validação por um especialista da área da saúde no domínio da Cardiopatia Isquêmica. Para facilitar esta validação, fez-se necessária a implementação da interface do módulo RBC, pois a versão atual estava implementada diretamente na base de dados, com difícil manipulação. Ao final, este módulo será utilizado pelo especialista médico para comparação dos resultados utilizando a definição de pesos por especialistas médicos com os resultados obtidos a partir da definição de pesos pelo AG.

Outra fase do trabalho foi a implementação do módulo AG conforme o modelo GAC-BRASIL, já explanado, no qual a função *fitness* é baseada na própria função de similaridade global do módulo RBC, aplicada em uma massa de dados para treinamento.

Para o melhor entendimento deste estudo, este capítulo vai mostrar, na seqüência: (i) a construção do módulo RBC; (ii) o ponto onde se aplicam pesos da função de similaridade global (objeto deste estudo); (iii) a construção do AG baseado no modelo GAC-BRASIL; (iv) as simulações para determinação dos pesos através do AG; e (v) a utilização destes pesos no módulo RBC.

#### **4.1. Implementação de interface para o módulo RBC**

O modelo RBC para o Projeto IACVIRTUAL (SILVA, 2005a) prevê que cada caso terá 17 sintomas ou características do paciente. O usuário (especialista médico) informa os dados do problema corrente, seleciona o conjunto de pesos que deseja utilizar e então o sistema:

- a) Seleciona o primeiro caso da base e calcula a similaridade local de cada uma das 17 características ou sintomas deste caso com o problema corrente;
- b) Calcula a similaridade global entre este caso e o problema corrente, aplicando o conjunto de pesos escolhido sobre os resultados das similaridades locais;
- c) Memoriza esta similaridade global e repete os passos anteriores para todos os casos da base;

- d) Ao final, ordena todos os casos em ordem decrescente de similaridade global e apresenta os casos com maior similaridade ao especialista médico para sua avaliação sobre a reutilização ou adaptação para o problema corrente da solução (tratamento) aplicada no caso da base.

#### 4.1.1. Problema corrente

O passo inicial, informação das características e sintomas do problema corrente, está ilustrado na tela da Figura 7.

Figura 7 - Tela de informações do problema corrente

#### 4.1.2. Similaridade local

A similaridade local será calculada para cada uma das 17 características ou sintomas do problema corrente. As características e as respectivas funções de similaridade local, previstas no modelo RBC do Projeto IACVIRTUAL (SILVA, 2005a), são mostradas na Tabela 3.

Tabela 3 - Características e sintomas dos casos

<b>Característica ou Sintoma do Paciente / Caso</b>	<b>Forma ou unidade de informação</b>	<b>Função de Similaridade Local</b>
Idade	informada em anos e então classificada em um das 18 faixas previstas	Função Linear
Sexo	M ou F	Função 0 ou 1
Pressão sistólica	informada em mmHg e então classificada em uma das 9 faixas previstas	Função Linear
Pressão diastólica	informada em mmHg e então classificada em uma das 9 faixas previstas	Função Linear
Frequência cardíaca	informada em bpm e então classificada em uma das 8 faixas previstas	Função Linear
Intensidade da dor	um código representando uma das 3 faixas previstas	Função Linear
Tempo de início do último episódio de dor	um código representando uma das 4 faixas previstas	Função Linear
Tempo de início dos sintomas	um código representando uma das 8 faixas previstas	Função Linear
Antecedentes	informação da existência ou não de 12 tipos de antecedentes	Função Intersecção Ponderada
Característica da dor	informação da existência ou não de 7 características específicas da dor	Função Intersecção Ponderada
Alterações no Eletrocardiograma	informação da existência ou não de 10 tipos de alterações no ECG	Função Intersecção Ponderada
Local da dor	informação da existência ou não de dor em 6 locais do corpo	Função Intersecção Ponderada
Outros sintomas	informação da existência ou não de 6 outros sintomas	Função Intersecção Ponderada
Sinais de ausculta	informação da existência ou não de 4 características observadas na ausculta	Função Intersecção Ponderada
Sinais observados	informação da existência ou não de 5 sinais observados no exame	Função Intersecção Ponderada
Fatores de risco	informação da existência ou não de 7 fatores considerados de risco	Função Intersecção Ponderada
Sinais de pulso	informação da existência ou não de 4 características nos sinais de pulso	Função Intersecção Ponderada

No total, têm-se três distintas funções de similaridade local, que, ainda de acordo com o modelo RBC do Projeto IACVIRTUAL, são calculadas da maneira descrita na Tabela 4.

**Tabela 4 - Funções de similaridade local**

<b>Função de Similaridade Local</b>	<b>Forma de cálculo</b>	<b>Fórmula</b>
Função 0 ou 1	Verifica se os elementos são iguais.	$v(v_i, v_j) = 1$ se $v_i = v_j$ ou $v(v_i, v_j) = 0$ se $v_i \neq v_j$
Função Linear	A similaridade cresce com o decréscimo da distância entre os dois valores, ponderada pelo tamanho do intervalo assumido pelo domínio.	$v(v_i, v_j) = 1$ se $v_i = v_j$ senão $v(v_i, v_j) = (1 - (v_i - v_j)) / (\text{limite superior} - \text{limite inferior})$
Função Intersecção Ponderada	A similaridade é proporcional aos elementos em comum dos dois casos, ponderada pelos elementos possíveis.	$v(v_i, v_j) = (\text{soma dos pesos dos elementos em comum entre } v_i \text{ e } v_j) / (\text{soma dos pesos de todos os elementos})$

As funções de similaridade local devem calcular as similaridades entre uma característica do problema corrente e a característica equivalente do caso da base. Por exemplo: no caso da característica “sexo do paciente”, que utiliza a Função 0 ou 1, cada caso da base que possuir o mesmo valor para a característica “sexo do paciente” contabiliza um para a similaridade local desta característica, enquanto os casos que possuem valores diferentes contabilizam zero. As demais características, que utilizam a Função Linear ou a Função Intersecção ponderada contabilizarão valores entre zero e um, de acordo com o grau de similaridade daquela característica. Estes valores serão utilizados no cálculo da similaridade global do caso.

#### **4.1.3. Similaridade global**

A função de similaridade global deve calcular a similaridade do problema corrente com o caso da base como um todo. A função utilizada pelo modelo RBC do Projeto IACVIRTUAL é a função do vizinho mais próximo. A fórmula desta função é a mesma descrita na equação 2 da seção “2.3. Cálculo de similaridade” deste trabalho. Esta função

utiliza pesos para cada uma das características, representando a relevância diferenciada de cada uma delas para a determinação de similaridade do caso com o problema corrente.

**Tabela 5 - Pesos do modelo atual**

<b>Ordem</b>	<b>Característica ou Sintoma do Paciente / Caso</b>	<b>Peso</b>
1	Sexo	0,6
2	Intensidade da dor	0,3
3	Tempo de início do último episódio de dor	0,4
4	Tempo de início dos sintomas	0,6
5	Frequência cardíaca	0,7
6	Pressão diastólica	0,7
7	Pressão sistólica	0,6
8	Idade	0,5
9	Alterações no Eletrocardiograma	1,0
10	Antecedentes	0,6
11	Característica da dor	0,8
12	Fatores de risco	0,7
13	Local da dor	1,0
14	Outros sintomas	0,5
15	Sinais de ausculta	0,7
16	Sinais observados	0,6
17	Sinais de pulso	0,4

O atual modelo utiliza os pesos apresentados na Tabela 5, atribuídos pelo Dr. Hervaldo Sampaio de Carvalho, especialista médico da área de Cardiopatia Isquêmica, e professor da Universidade de Brasília (UnB) com atuação em Informática Médica e Cardiologia Clínica.

Estes pesos foram determinados pelo especialista com base na literatura publicada e em sua experiência clínica nesta área da cardiologia.

## **4.2. Implementação do módulo AG**

De acordo com o que já foi apresentado, este estudo pretende propor o modelo GAC-BRASIL, que foi detalhado no capítulo de metodologia, e implementar um módulo baseado neste modelo. O intuito é propor um conjunto de pesos, fundamentados em uma técnica de aprendizado de máquina utilizando bases de casos reais para auxiliar os



especialistas médicos, usuários do módulo RBC, na definição de pesos para cada uma das características e/ou sintomas dos casos para determinação de sua similaridade global com o problema corrente.

#### **4.2.1. Preparação dos dados para a função *fitness***

A atual base do módulo RBC está em formato *Microsoft Access*. Assim, foi utilizado este mesmo formato para a preparação para o módulo AG.

O modelo GAC-BRASIL determina que a função de aptidão (*fitness*) do AG seja a mesma função de similaridade do modelo RBC. Desta forma são necessários problemas correntes e uma base de casos para comparação com os problemas correntes, sendo que se deve conhecer o diagnóstico real de ambos. Nesta implementação, estes dois grupos de dados para treinamento do AG serão formados a partir de uma extração parcial da própria base de dados do módulo RBC. Nesta base, todos os casos possuem um dos dois tipos de diagnósticos: diagnóstico real de Infarto Agudo do Miocárdio (IAM) ou possuem diagnóstico real de angina estável (baixa probabilidade de IAM).

Dentre os 529 casos da base, alguns não possuem as informações de todas as 17 características ou sintomas. Embora tenham a “qualidade necessária” (SILVA, 2005a, p. 94) para o uso em RBC, para efeito de calibração dos pesos através do AG, que é o objetivo deste trabalho, faz-se necessário trabalhar com casos que possuam o maior número de informações, pois os parâmetros de similaridade local não presentes em dois casos comparados são desconsiderados para o cálculo da similaridade global.

Apenas 8 dos casos possuem todos os 17 parâmetros, e 44 têm apenas uma omissão, possuindo, portanto, 16 parâmetros para comparação. Em princípio, os dois grupos de casos já explicados na metodologia (grupo menor de supostos problemas correntes e grupo maior de casos para comparação) foram compostos respectivamente por estes dois conjuntos com 8 e 44 casos, respectivamente.

Contudo, nas primeiras simulações, as notas de aptidão eram muito similares entre os indivíduos, e, independentemente das alterações efetuadas nos demais parâmetros como o percentual de mutação, cruzamento, e forma de seleção, elas sempre se mantinham num percentual próximo.

Observou-se que no primeiro grupo, 7 dos 8 casos possuíam o mesmo diagnóstico (alta probabilidade de IAM), o mesmo acontecendo no segundo grupo, com 34 dos 44 casos também possuíam este diagnóstico. Optou-se então por recriar os grupos, desta vez com mais casos e com uma composição equilibrada. Os novos casos acrescentados possuíam informações para 15 das 17 características ou sintomas.

O primeiro grupo ficou com a seguinte composição:

**Tabela 6 - Composição do grupo de “problemas correntes” da função *fitness***

Total de casos	Distribuição por Diagnóstico		Total de características/sintomas registrados	
	16 casos	8 casos	alta probabilidade de IAM	7 casos
1 caso				16 informações
8 casos		baixa probabilidade de IAM	1 caso	17 informações
			7 casos	16 informações

O segundo grupo ficou com 64 casos, dispostos da seguinte maneira:

**Tabela 7 - Composição do grupo de “casos da base” da função *fitness***

Total de casos	Distribuição por Diagnóstico		Total de características/sintomas registrados	
	64 casos	32 casos	alta probabilidade de IAM	32 casos
04 casos				16 informações
32 casos		baixa probabilidade de IAM	28 casos	15 informações

#### 4.2.2. Configuração dos parâmetros do AG

O modelo GAC-BRASIL indica que os parâmetros do AG devem ser configuráveis, permitindo a seleção das melhores opções, de acordo com o desempenho e resultados.

Para isto, foi criada a tela apresentada na Figura 8. Nesta tela, caso seja selecionada a opção de representação em cadeia de *bits* (“0 - Binária”), o tipo de cruzamento permitido é o *crossover* de um ponto e o tipo de mutação é substituição binária. Caso seja escolhida a opção de representação por números reais, o tipo de cruzamento permitido é o *crossover* média e o tipo de mutação é a substituição de genes por outro valor.

Parâmetros do Algoritmo Genético	
Representação	0 - Binária
Tamanho População	72
Qtde Genes p/ Indiv.	17
Bits por Gene	4
% de Cruzamento	75
Tipo de Cruzamento	0 - Crossover de um ponto
% de Mutação	3
Tipo de Mutação	0 - Substituição binária
Tipo de Seleção	0 - Roleta Ponderada
Elitismo	1 - Garante melhor indivíduo na população
Critério de Parada	1 - Quando atingir número máximo de gerações
Máx. de Gerações	25
% Variação Fitness	1
% Limiar Similaridade	70

**Figura 8 - Tela de parametrização do AG**

O parâmetro de tamanho da população indica quantos indivíduos serão criados a cada geração. Quanto maior o número de indivíduos, maior será a cobertura do espaço de busca. O limitador para este parâmetro é o tempo de processamento. Foram testados diversos tamanhos de população e, devido à rápida resposta do algoritmo, este número foi sendo gradativamente incrementado até que se optou por ficar com o número de 150 membros por geração,

quantidade de indivíduos que, pelos resultados das simulações, ainda apresentou uma performance satisfatória, enquanto que quantidades superiores começaram a degradar o desempenho sem apresentar diferença significativa no ponto de convergência.

A quantidade de genes por indivíduo, neste contexto, deve ser igual a 17, pois se tem esta mesma quantidade de sintomas/características, e, de acordo com o modelo GAC-BRASIL, cada gene deve representar um peso da função de similaridade global do RBC.

A quantidade de *bits* por gene vai determinar os valores mínimos e máximos para cada gene. Com quatro *bits* por gene, em representação binária o valor mínimo é “0000” e o valor máximo é “1111”. Em representação por números reais, os valores máximo e mínimo são, respectivamente, 0 e 15. Estes valores não interferem no cálculo de similaridade, mas, no final, para se utilizar os valores dos genes no módulo RBC, os valores devem ser normalizados entre 0 e 1, dividindo-se cada valor por 15.

Os percentuais de cruzamento e mutação e a definição de elitismo interferem diretamente na diversidade da população e na convergência das gerações. Enquanto que, por um lado, as alterações genéticas garantem a diversidade da população, por outro, podem causar a perda de bons membros. O fenômeno da convergência ocorre quando se verifica pouca ou nenhuma variação do *fitness* médio da população de uma geração para outra, indicando pouca diversidade na população. Após algumas simulações com percentuais de 70% para cruzamento, 1% para mutação e sem a aplicação de elitismo, verificou-se que uma convergência prematura, por volta da 20ª geração. Desta forma, embora dentro da faixa sugerida pela literatura pesquisada, optou-se pela escolha de valores altos para os percentuais de probabilidade de cruzamento e mutação, respectivamente 80% e 3%, e, para compensar esta alta diversidade, foi implementado o elitismo, que garante o membro com maior *fitness* na próxima geração.

A forma de seleção de indivíduos é outro fator que contribui para a variação do *fitness* médio entre cada geração. O modelo GAC-BRASIL sugere o *ranking*. Quando se trata de casos em que a nota de *fitness* é muito próxima, ou com extremos, o *ranking* é mais seletivo do que a roleta. Esta diferença pode ser observada na Tabela 8, onde são comparadas as probabilidades de seleção dos mesmos casos, com a mesma nota, nos dois métodos. Enquanto que numa seleção por roleta o sorteio é feito com base na soma das notas de *fitness*, no *ranking* o sorteio é feito com base na soma das posições de cada caso. Neste exemplo, a melhor solução (caso *a*) teria 22,9% de chance de ser selecionado na roleta e 33,3% de chance de ser selecionado no *ranking*. Já as probabilidades da pior solução (caso *e*) seriam 17,1% na roleta e apenas 6,7% no *ranking*.

**Tabela 8 - Comparação roleta x *ranking***

Caso	Roleta - sorteio entre 1 e 350			<i>Ranking</i> - sorteio entre 1 e 5		
	<i>Fitness</i>	Faixa para seleção	Probabilidade	Ordem	Faixa para seleção	Probabilidade
<i>a</i>	80	1 a 80	22,9%	5	1 a 5	33,3%
<i>b</i>	75	81 a 155	21,4%	4	6 a 9	26,7%
<i>c</i>	70	156 a 225	20,0%	3	10 a 12	20,0%
<i>d</i>	65	226 a 290	18,6%	2	13 a 14	13,3%
<i>e</i>	60	291 a 350	17,1%	1	15 a 15	6,7%
<b>Total:</b>	<b>350</b>		<b>100,0%</b>	<b>15</b>		<b>100,0%</b>

O critério de parada do AG pode ser em relação à convergência, ao número de gerações, ou ainda uma combinação destes dois parâmetros. Neste último caso, utilizado nesta implementação, o AG só vai parar a evolução quando a variação do *fitness* médio de uma geração para outra for inferior a 1%, mas não antes de ser atingida a 50ª geração. Esta opção foi escolhida por reservar a possibilidade de que após um período de convergência (estabilidade), uma mutação genética possa alterar o equilíbrio e gerar um indivíduo acima da média.

O percentual de similaridade é o limiar utilizado para qualificar, durante a função *fitness*, se o caso é ou não similar. Isto quer dizer que após calcular a similaridade global do

problema corrente com o caso da base, o diagnóstico deste só é comparado e considerado para efeito do cálculo da nota, caso a similaridade global tenha atingido o limiar. Foram feitos testes com valores entre 50% e 90% para este parâmetro. Com os valores mais baixos, observou-se pouca variação entre as notas. Com os valores mais altos a variação aumentou, contudo, observaram-se vários casos com nota zero, e a convergência estava ocorrendo prematuramente, antes da décima geração. O limiar utilizado foi de 70%.

### 4.3. Representação por dígitos binários x números reais

Comparando os resultados dos algoritmos executados com os cromossomos representados por dígitos binários com aqueles executados com representação por números reais, mantendo todos os demais parâmetros com as mesmas configurações, os principais pontos observados foram as seguintes:

- Devido à geração aleatória da primeira geração, nos dois tipos de representação foi obtido um *fitness* médio desta primeira geração na faixa de 61%.
- Nas execuções com números reais, o *fitness* médio das quinquagésimas (últimas) gerações variou entre 71% e 77%. Estes valores foram nitidamente inferiores aos obtidos nas execuções com números binários, nas quais o *fitness* médio das últimas gerações variou entre 88% e 92%.
- O mesmo aconteceu com o melhor indivíduo de cada algoritmo. Nas execuções em que o cromossomo tinha representação por número real, o *fitness* da melhor solução variou entre 76,4% e 78,1%, enquanto que nas execuções em que o cromossomo tinha representação por dígitos binários, o *fitness* da melhor solução variou entre 88,8% e 92,8%.
- Outra diferença observada, e que pode explicar os resultados já apresentados foi com relação à convergência. Com números reais, o *fitness* médio da 15ª geração

ficou menos de um ponto percentual abaixo do *fitness* médio da 50ª, sendo que em alguns casos isto foi observado já a partir da 10ª geração. Contudo, com dígitos binários, esta situação só foi observada a partir da 25ª geração.

A diferença entre o AG com indivíduos representados por números reais e o AG com indivíduos representados por dígitos binários, com relação à evolução do *fitness* médio e do *fitness* do melhor indivíduo, destacando a linha de convergência, está ilustrada nos gráficos apresentados na Figura 9 (números reais) e Figura 10 (dígitos binários).

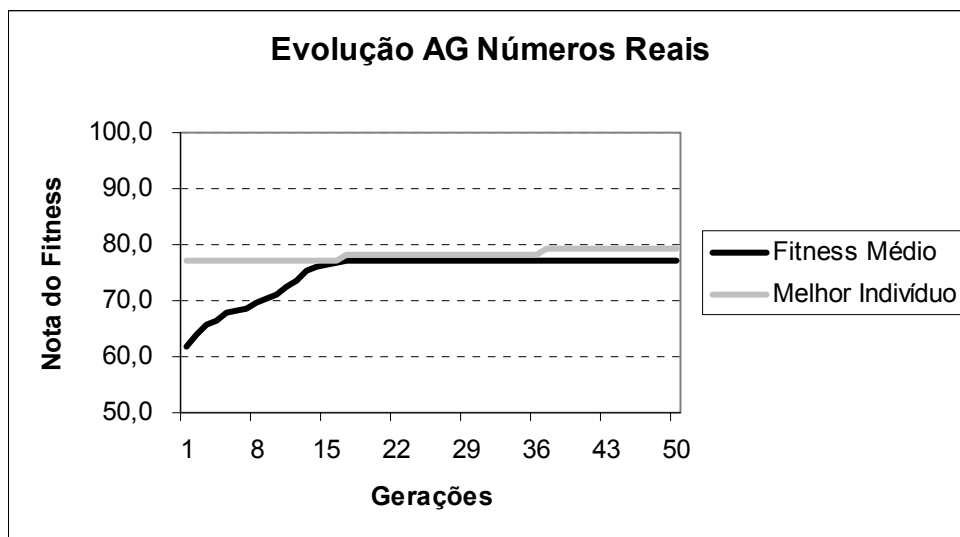


Figura 9 - Gráfico de evolução do *fitness* em AG com números reais

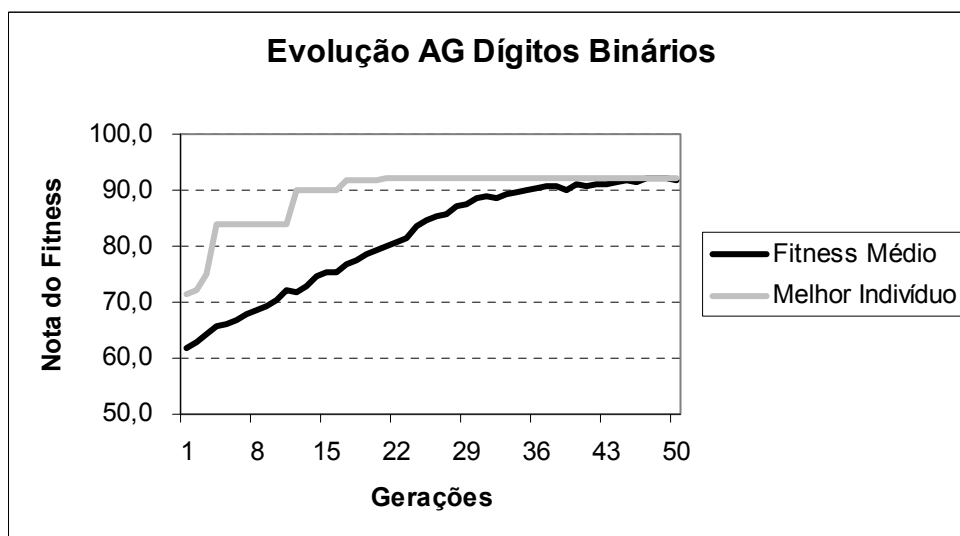


Figura 10 - Gráfico de evolução do *fitness* em AG com dígitos binários

Também foram encontradas diferenças entre os genes das melhores soluções de cada tipo de representação. A Tabela 9 apresenta os valores dos genes da melhor solução de cinco execuções do AG, em cada um dos dois tipos de representação, sendo que, conforme explicado anteriormente, os valores possíveis para os pesos da função de similaridade global são entre 0 e 1, e os genes de cada indivíduo representam cada um dos 17 pesos desta função.

**Tabela 9 - Valores médios dos genes das melhores soluções do AG**

AG Representação Dígitos Binários							AG Representação Números Reais						
Pesos	A	B	C	D	E	Média	Pesos	A	B	C	D	E	Média
1	0,47	0,27	0,47	0,07	0,07	<b>0,27</b>	1	0,38	0,26	0,35	0,27	0,66	<b>0,38</b>
2	0,47	0,47	0,33	0,67	1,00	<b>0,59</b>	2	0,61	0,49	0,38	0,43	0,64	<b>0,51</b>
3	0,20	0,13	0,27	0,27	0,47	<b>0,27</b>	3	0,43	0,28	0,51	0,26	0,33	<b>0,36</b>
4	0,53	0,80	0,07	0,07	0,33	<b>0,36</b>	4	0,20	0,34	0,28	0,20	0,62	<b>0,33</b>
5	0,60	0,53	0,87	0,07	0,20	<b>0,45</b>	5	0,43	0,28	0,53	0,33	0,49	<b>0,41</b>
6	0,40	0,07	0,47	0,93	0,07	<b>0,39</b>	6	0,46	0,46	0,58	0,13	0,17	<b>0,36</b>
7	0,20	0,33	0,93	0,07	1,00	<b>0,51</b>	7	0,72	0,80	0,47	0,25	0,46	<b>0,54</b>
8	0,27	0,07	0,20	0,13	0,33	<b>0,20</b>	8	0,32	0,43	0,52	0,49	0,41	<b>0,43</b>
9	0,93	0,87	1,00	0,67	0,93	<b>0,88</b>	9	0,78	0,67	0,82	0,83	0,53	<b>0,73</b>
10	1,00	0,53	0,73	1,00	0,80	<b>0,81</b>	10	0,68	0,82	0,74	0,81	0,63	<b>0,74</b>
11	0,20	0,33	0,20	0,53	0,87	<b>0,43</b>	11	0,43	0,33	0,25	0,85	0,50	<b>0,47</b>
12	0,93	0,60	0,93	0,47	0,80	<b>0,75</b>	12	0,71	0,88	0,83	0,80	0,57	<b>0,76</b>
13	0,87	0,87	0,67	0,87	1,00	<b>0,85</b>	13	0,77	0,82	0,79	0,94	0,80	<b>0,82</b>
14	0,87	0,73	0,80	0,87	0,47	<b>0,75</b>	14	0,64	0,53	0,53	0,47	0,70	<b>0,57</b>
15	0,33	0,40	0,80	0,13	0,07	<b>0,35</b>	15	0,64	0,16	0,45	0,29	0,35	<b>0,38</b>
16	0,67	0,80	0,73	0,73	0,87	<b>0,76</b>	16	0,71	0,44	0,60	0,52	0,49	<b>0,55</b>
17	0,20	0,07	0,13	0,07	0,07	<b>0,11</b>	17	0,64	0,31	0,22	0,14	0,53	<b>0,37</b>

Esta média, calculada a partir dos genes do melhor indivíduo de cinco execuções de cada tipo de AG foi transferida para o módulo RBC, conforme é descrito adiante.

#### 4.4. Definição de pesos por um grupo de especialistas médicos

Além da definição de pesos pelo AG, também foi coletada a opinião de especialistas médicos, para determinar um outro conjunto de pesos, baseados no seu conhecimento e experiência clínica.



Foi distribuído um formulário (Figura 11) para especialistas da área de Cardiologia, solicitando que atribuísem, de acordo com o grau de importância para o diagnóstico de Cardiopatia Isquêmica Aguda (Infarto Agudo do Miocárdio), pesos entre 1 e 10 para os 17 sintomas/características que compõem este problema.

Caro especialista médico,  
 Dentro de um projeto de pesquisa da área de Informática na Saúde, temos um banco de dados contendo casos reais de Cardiopatia Isquêmica.  
 Nosso projeto consiste basicamente em, dada uma nova situação, buscar os casos mais similares desta base histórica, para efeito de suporte ao diagnóstico médico.  
 Para isto, precisamos identificar qual o peso (relevância) de cada um dos 17 atributos registrados na base.  
 Pedimos que preencha, com **valores inteiros entre 1 e 10**, o peso que considera ideal para cada um dos atributos abaixo:

Nome (opcional): \_\_\_\_\_

Experiência em Cardiologia (em anos/meses): \_\_\_\_\_

Atributo	Peso Proposto
Idade do paciente	
Sexo do paciente	
Pressão sistólica do paciente	
Pressão diastólica do paciente	
Frequência cardíaca do paciente	
Tempo decorrido desde o último episódio de dor	
Tempo decorrido desde o início dos sintomas de dor	
Intensidade da dor	
Característica da dor	
Local da dor	
Alterações do ECG (BRE, SVE, marcapasso, etc.).	
Antecedentes (intervenções coronárias, histórico de doenças, etc.).	
Fatores de risco (diabete, tabagismo, obesidade, etc.).	
Sinais de ausculta	
Sinais de pulso (arritmico, filiforme, etc.).	
Sinais observados (palidez, reação à palpação, vômitos, etc.).	
Outros sintomas (febre, sudorese, náuseas, etc.).	

**Figura 11 - Formulário para indicação de pesos por especialistas**

O formulário foi respondido por cinco especialistas médicos com uma média de 13 anos de experiência clínica em cardiologia. Este grupo, dentre outras atividades, faz parte do corpo médico do Hospital Universitário da Universidade de Brasília (HUB).

O resultado desta pesquisa, com a média dos pesos atribuídos já normalizados para o padrão entre 0 e 1 é apresentado na Tabela 10.

**Tabela 10 - Média dos pesos atribuídos pelo grupo de especialistas**

Ordem	Característica ou Sintoma do Paciente / Caso	Pesos por especialista					Média
		A	B	C	D	E	
1	Sexo	0,7	0,8	0,8	0,9	0,5	0,74
2	Intensidade da dor	0,8	0,7	1	1	1	0,90
3	Tempo de início do último episódio de dor	1	0,8	1	0,9	0,9	0,92
4	Tempo de início dos sintomas	1	1	1	0,9	1	0,98
5	Frequência cardíaca	0,5	0,5	0,8	0,8	0,7	0,66
6	Pressão diastólica	1	0,7	0,8	0,9	0,8	0,84
7	Pressão sistólica	1	0,7	0,8	0,9	0,9	0,86
8	Idade	1	0,8	0,8	1	0,9	0,90
9	Alterações no Eletrocardiograma	1	0,8	1	0,9	1	0,94
10	Antecedentes	1	0,8	1	1	1	0,96
11	Característica da dor	0,8	1	1	1	1	0,96
12	Fatores de risco	1	0,9	1	0,8	1	0,94
13	Local da dor	0,8	1	1	0,9	0,9	0,92
14	Outros sintomas	0,5	0,7	1	0,7	0,8	0,74
15	Sinais de ausculta	0,6	0,6	1	0,7	1	0,78
16	Sinais observados	0,7	0,8	1	0,9	0,9	0,86
17	Sinais de pulso	0,6	0,9	1	0,8	0,9	0,84
<b>Tempo de experiência em Cardiologia (em anos):</b>		<b>7</b>	<b>20</b>	<b>-x-</b>	<b>15</b>	<b>10</b>	<b>13</b>

## 4.5. Utilização dos pesos no módulo RBC

### 4.5.1. Escolha do grupo de pesos

No módulo RBC, após o especialista médico preencher os campos da tela do problema corrente apresentada na Figura 7, ele deve ter a opção de escolha do conjunto de pesos a ser utilizado.

Os quatro grupos de pesos previstos são: o definido pelo especialista no modelo RBC de 2005 (Tabela 5); a média dos pesos atribuídos pelo grupo de especialistas em 2006 (Tabela 10); e os dois resultados médios das execuções do AG com dígitos binários e números reais (Tabela 9). Estas informações podem ser comparadas na Tabela 11 e no gráfico a seguir.

Tabela 11 - Comparação dos pesos atribuídos pelos especialistas e pelo AG

Seq.	Característica / Sintoma	Especialista (2005)	Grupo Especialistas (2006)	AG dígitos binários	AG números reais
1	Sexo	0,6	0,74	0,27	0,38
2	Intensidade da dor	0,3	0,9	0,59	0,51
3	Tempo de início do último episódio de dor	0,4	0,92	0,27	0,36
4	Tempo de início dos sintomas	0,6	0,98	0,36	0,33
5	Frequência cardíaca	0,7	0,66	0,45	0,41
6	Pressão diastólica	0,7	0,84	0,39	0,36
7	Pressão sistólica	0,6	0,86	0,51	0,54
8	Idade	0,5	0,9	0,2	0,43
9	Alterações no Eletrocardiograma	1	0,94	0,88	0,73
10	Antecedentes	0,6	0,96	0,81	0,74
11	Característica da dor	0,8	0,96	0,43	0,47
12	Fatores de risco	0,7	0,94	0,75	0,76
13	Local da dor	1	0,92	0,85	0,82
14	Outros sintomas	0,5	0,74	0,75	0,57
15	Sinais de ausculta	0,7	0,78	0,35	0,38
16	Sinais observados	0,6	0,86	0,76	0,55
17	Sinais de pulso	0,4	0,84	0,11	0,37

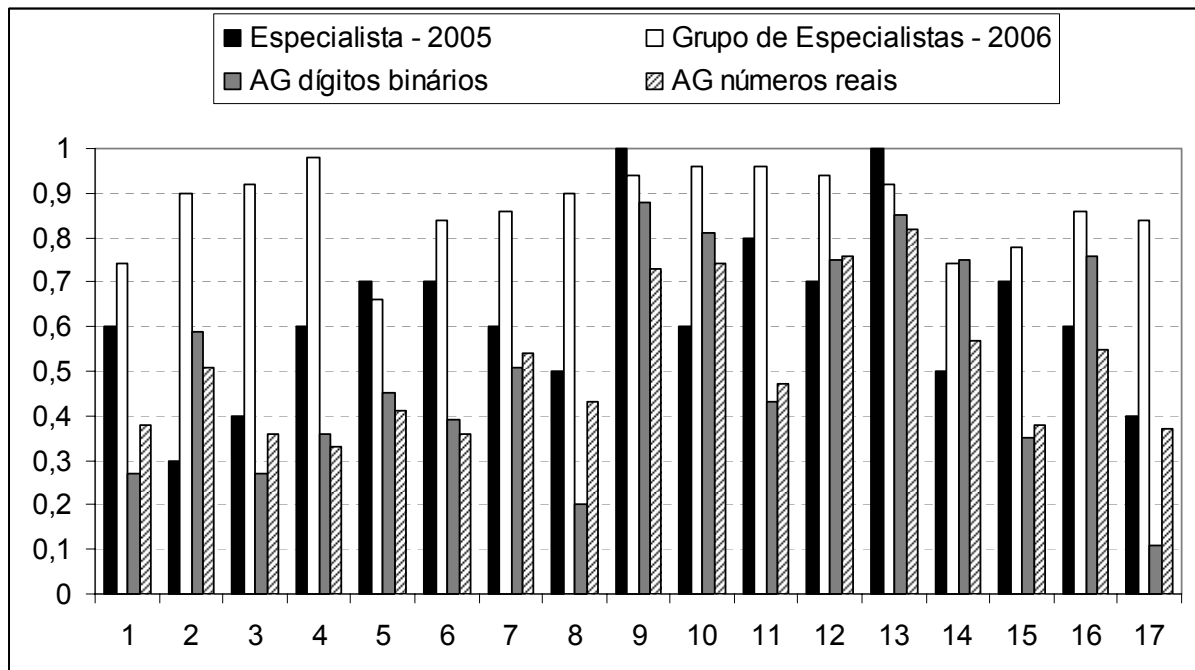


Figura 12 - Gráfico comparativo entre pesos atribuídos pelos especialistas e pelo AG

Então, o sistema apresenta a tela ilustrada na Figura 13, mostrando os quatro conjuntos de pesos e permitindo a escolha daquele conjunto que será utilizado para o cálculo da similaridade do problema corrente com os casos da base.

Após esta escolha, o módulo RBC calcula a similaridade local entre as características do problema corrente e os casos da base, e em seguida a similaridade global, aplicando os pesos escolhidos sobre os resultados dos cálculos locais.

**Seleciona Conjunto de Pesos para Cálculo da Similaridade**

Característica	Especialista	Grupo	AG Bin.	AG Real
Sexo do Paciente	0,6	0,74	0,27	0,38
Intensidade da Dor	0,3	0,9	0,59	0,51
Tempo do Últ. Episódio de Dor	0,4	0,92	0,27	0,36
Tempo do Início dos Sintomas	0,6	0,98	0,36	0,33
Frequência Cardíaca	0,7	0,66	0,45	0,41
Pressão Diastólica	0,7	0,84	0,39	0,36
Pressão Sistólica	0,6	0,86	0,51	0,54
Idade do Paciente	0,5	0,9	0,2	0,43
Alterações do ECG	1	0,94	0,88	0,73
Antecedentes	0,6	0,96	0,81	0,74
Característica da Dor	0,8	0,96	0,43	0,47
Fatores de Risco	0,7	0,94	0,75	0,76
Local da Dor	1	0,92	0,85	0,82
Outros Sintomas	0,5	0,74	0,75	0,57
Sinais de Ausculta	0,7	0,78	0,35	0,38
Sinais Observados	0,6	0,86	0,76	0,55
Sinais de Pulso	0,4	0,84	0,11	0,37

Conjunto de Pesos a ser utilizado para busca:

1 - Pesos do Especialista (2005)  
 2 - Pesos do Grupo de Especialistas (2006)  
 3 - Pesos do AG com dígitos binários  
 4 - Pesos do AG com números reais

**Figura 13 - Tela de escolha do conjunto de pesos**

#### 4.5.2. Apresentação dos resultados do RBC

Todos os casos são ordenados em ordem decrescente de similaridade global e os 10 casos mais similares, juntamente com o problema corrente, são apresentados ao usuário, conforme a Figura 14.

Nesta tela, são listadas as seguintes informações:

- Número do caso (registro na base);
- Percentual de similaridade com o problema corrente;

- Código do diagnóstico real (1-baixa probabilidade de IAM, 3-alta probabilidade de IAM);
- Os dados das 17 características e seus detalhes.

Resultado da Busca											
Característica	Problema	r.01	r.02	r.03	r.04	r.05	r.06	r.07	r.08	r.09	r.10
Caso Número:		3240	2888	3113	2887	2997	2686	3025	3162	2942	3191
% Similaridade:		81,79	80,17	79,78	73,90	73,73	73,73	73,43	73,42	73,09	72,08
Diagnóstico Real:		3	3	3	3	3	3	3	3	3	3
Idade do Paciente	65	51	58	54	66	62	67	58	53	53	51
Sexo do Paciente: (M)asculino / (F)Feminino	M	M	M	M	M	M	F	M	M	M	F
Quadro Dor - Intensidade da dor do paciente (1 a 3)	3	3	3	2	2	3	3	2	2	3	3
Quadro Dor - Cód. tempo início do último episódio de dor	0	1	1	0	1	1	0	0	0	1	0
Quadro Dor - Cód. tempo início dos sintomas de dor - N	0	2	0	0	7	6	6	3	7	0	2
Exame Básico - Frequência Cardíaca do paciente (bpm)	88	96	104	66	90	60	90	80	80	84	88
Exame Básico - Pressão Diastólica do paciente (mmHg)	90	100	100	90	90	70	100	80	80	100	100
Exame Básico - Pressão Sistólica do paciente (mmHg)	160	140	170	140	160	160	160	140	150	200	170
Característica da Dor 01 - Dependente da posição do p											
Característica da Dor 02 - Desencadeada com esforços						X					
Característica da Dor 03 - Em aperto ou opressão	X	X		X	X	X	X	X	X		X
Característica da Dor 04 - Em peso ou queimor			X			X					
Característica da Dor 05 - Em pontada						X					
Característica da Dor 06 - Espontânea em repouso	X	X	X	X	X	X	X	X	X	X	
Característica da Dor 07 - Ventilatório dependente											
Local da Dor 01 - Braço esquerdo e/ou direito								X			
Local da Dor 02 - Dorso			X								
Local da Dor 03 - Hemitórax direito			X								
Local da Dor 04 - Mandíbula						X					
Local da Dor 05 - Região epigástrica											
Local da Dor 06 - Região precordial ou médio-esternal	X	X	X	X	X	X	X	X	X	X	X
Alteração ECG 01 - Marcapasso											
Alteração ECG 02 - BRE											
Alteração ECG 03 - SVE											
Alteração ECG 04 - Segmento ST supradesnivelado	X	X	X							X	
Alteração ECG 05 - Segmento ST infradesnivelado											
Alteração ECG 06 - Onda T Apiculada											

Figura 14 - Tela de apresentação dos resultados do RBC

O objetivo desta tela é apresentar ao especialista os casos que o RBC recuperou, permitindo a utilização e/ou adaptação de uma ou mais soluções para o problema corrente.

Com este módulo desenvolvido foi possível a validação dos resultados do AG.

#### 4.6. Simulações do RBC para avaliação dos resultados do AG

Após as simulações do AG, faz-se necessária a validação dos resultados. Ou seja, avaliar quanto pode ser útil para o especialista a sugestão de pesos do AG.

Para isto, contou-se com o auxílio do mesmo especialista médico que validou o modelo RBC (SILVA, 2005a), Dr. Hervaldo Sampaio Carvalho.

#### **4.6.1. Primeiras simulações e ajuste na base de dados**

Foram montados, com base na experiência do especialista, cenários de baixa probabilidade de IAM e cenários de alta probabilidade de IAM.

Estes cenários foram aplicados e os primeiros resultados não se mostraram satisfatórios. Para cenários considerados de claro indício de IAM, independentemente dos conjuntos de pesos utilizados, dentre os dez casos mais similares foram retornados de três a cinco casos de baixa probabilidade.

Para se chegar às causas desta divergência, foram analisados os dados detalhados dos casos retornados, constatando-se que os casos divergentes possuíam muitas lacunas de informações. Conforme o modelo RBC do Projeto IACVIRTUAL (SILVA, 2005a, ALMEIDA, 2006a), as características ausentes no caso da base não devem ser consideradas para o cálculo da similaridade global. Ao desconsiderar uma similaridade local no cálculo global, o seu peso também não é considerado, fazendo com que casos com muitas lacunas e poucas características/sintomas em comum fossem considerados bastante similares, levando ao erro de diagnóstico.

O levantamento dos casos da base mostrou que existiam casos com informações para apenas 11 ou 12 das 17 características previstas no modelo. O modelo RBC foi baseado na metodologia *Cross-industry Standard Process for Data Mining* (CRISP-DM), cujo ciclo prevê em sua fase de Modelagem dos Dados, o retorno à fase de Preparação dos Dados (SILVA, 2005a; LAROSE, 2005).

Assim, foi refeita a preparação dos dados, retirando-se da base todos os casos com menos de 14 das 17 informações previstas. Restaram ainda 303 casos reais, o que foi considerado uma quantidade satisfatória pelo especialista.

#### 4.6.2. Simulações de RBC

Com a limpeza dos dados, foram novamente executadas simulações. Foi elaborado um formulário com todas as características consideradas no problema corrente. Solicitou-se ao especialista médico e sua equipe o preenchimento de casos diferentes entre si, mas que todos indicassem uma alta probabilidade de IAM. O mesmo foi feito para outro grupo de casos, com baixa probabilidade de IAM.

Os problemas de cada grupo de casos correntes foram testados no módulo RBC com cada um dos grupos de pesos. A média de casos retornados com diagnóstico igual ao esperado para cada grupo de pesos encontra-se na Tabela 12.

**Tabela 12 - Média de acertos (mesmo diagnóstico) por tipo de peso**

<b>Média de Casos Retornados com Diagnóstico Idêntico</b>			
<b>Especialista</b>	<b>Grupo de Especialistas</b>	<b>AG com Dígitos Binários</b>	<b>AG com Números Reais</b>
8,5	8,5	9	9

Os novos resultados foram considerados bastante satisfatórios pelo especialista, porque em todas as simulações a grande maioria dos casos retornados possuía o mesmo diagnóstico esperado para o problema corrente.

Embora os resultados de todos os conjuntos de pesos tenham sido de acordo com a expectativa, os resultados obtidos com os conjuntos de pesos retornados pelo AG obtiveram um índice de diagnósticos coincidentes da ordem de 90%, ligeiramente superior ao índice dos especialistas de 85%.

Contudo, mais importante para o especialista foi o fato de que os casos mais similares retornados pelas buscas feitas com os pesos dados pelo AG, eram, na maior parte das vezes, os mesmos casos retornados pelas buscas feitas com pesos atribuídos por especialistas. Isto demonstra que houve aprendizagem de máquina, haja vista que o AG, partindo de valores de

pesos totalmente aleatórios, sugeriu um conjunto de valores que retornou resultados coincidentes com os conjuntos de valores sugeridos pelos especialistas médicos.



## **5. Conclusões e Trabalhos Futuros**

## 5.1. Conclusões

Os resultados apresentados mostraram que o modelo GAC-BRASIL, que determina que a função *fitness* deve ser igual à função de similaridade do RBC e que devem ser utilizados casos com diagnósticos reais conhecidos para calibração do AG, apresentou uma solução satisfatória, haja vista que, partindo de valores totalmente aleatórios, chegou a um conjunto de pesos que apresentou desempenho similar ao conjunto de pesos apresentados por especialistas médicos, demonstrando que houve Aprendizado de Máquina. Isto pode ser comprovado não somente pelo fato dos conjuntos de pesos sugeridos pelo AG retornarem, para cada problema corrente, 90% de casos com o mesmo diagnóstico, mas, principalmente, porque na grande maioria das simulações os casos mais similares eram os mesmos retornados pelos conjuntos de pesos sugeridos por especialistas médicos.

A validação do módulo AG pelo especialista médico, comprova que o modelo GAC-BRASIL é adequado para auxiliar o especialista médico na determinação de pesos (relevância) de características e sintomas para o diagnóstico de Cardiopatia Isquêmica no contexto de RBC.

A otimização da ferramenta de busca e recuperação de casos do módulo RBC do Projeto IACVIRTUAL foi outro objetivo alcançado por este trabalho. O tempo de resposta, a forma de lançamento de informações e a apresentação dos casos foram avaliados satisfatoriamente pelo especialista médico.

Esta otimização foi importante não só para validar o método proposto, mas também para uma pequena revisão de uma fase do modelo RBC, trabalho que deu origem a esta pesquisa. Esta ação, embora não fosse objetivo original deste trabalho, foi importante para o Projeto IACVIRTUAL, haja vista que foi refeita a seleção dos casos a serem utilizados, numa revisão da fase de preparação dos dados da metodologia CRISP-DM. Esta constatação mostra

a importância da fase de preparação de dados para o modelo RBC, onde devem ser mantidos os casos com o maior número possível de informações, a fim de evitar distorções nos casos retornados como similares no processo de busca.

Outra conclusão que se tira do presente estudo, é que os AG com indivíduos representados por cadeias de dígitos binários apresentam, pelo menos neste caso particular de similaridade, soluções com melhor *fitness*. Os AG com indivíduos representados por vetores de números reais apresentaram uma convergência prematura, ficando, em todas as simulações, com uma pior curva de crescimento do *fitness* médio.

## 5.2. Trabalhos futuros

Como um dos trabalhos futuros pode-se indicar a especificação de um AG para determinação dos pesos das funções de similaridade local, o qual poderia auxiliar o especialista médico a definir a relevância de cada item do conjunto que compõe a característica ou sintoma.

Também como estudo futuro deve ser considerada a aplicação e validação do modelo GAC-BRASIL em outros modelos de RBC.

Outra investigação importante a ser realizada é a análise da causa da diferença entre os pesos atribuídos pelo especialista médico e pelo AG. Embora com resultados satisfatórios, alguns dos pesos apresentados pelo AG foram muito diferentes do que os especialistas médicos esperavam. Acredita-se que esta diferença pode ter sido causada pela característica dos casos da base, mas este é um ponto que deve ser objeto de uma pesquisa mais aprofundada.

Um trabalho, com alguns passos já em andamento, é a implementação de novas formas de seleção, por exemplo, a seleção por torneio, e novas análises de convergência da população.

Uma outra possibilidade, que merece um estudo, seria a criação de um AG que definisse os melhores parâmetros para o próprio AG dentro de um determinado contexto.

Esta dissertação complementou a construção do módulo RBC do Projeto IACVIRTUAL e com os resultados obtidos poderá ser uma ferramenta que venha a ser utilizada pela comunidade da área da saúde. Da mesma maneira, o modelo GAC-BRASIL, complementado por trabalhos futuros, poderá contribuir para a pesquisa na área de Inteligência Artificial e Aprendizado de Máquina.

## **Referências**

---

AAMODT, A., PLAZA, E. **Case Based Reasoning: Foundational Issues, Methodological variations and System Approaches**. Proceedings of the AI Communications, Vol. 7, N. 1, p.39-59, 1994.

ALGARVE, A.S. **Algoritmos Genéticos - Revisão da Teoria e Exemplo Didático Simples**. GPEB - UFSC, 1994.

ALMEIDA, C.W.D. et al. **Virtual Medical Office: A Medical and Educational Tool**. Proceedings of International Congress on Computation Bioengineering, Espanha, 2003b.

ALMEIDA, C.W.D. et al. **Virtual Web System for Realization of Consultation and Diagnosis**. Proceedings of the Conference on Medical Cybernetics in Clinical Practice in Burdenko Main Military Clinical Hospital, vol. 2, p.98-110, Moscow, Russia, 2004a.

ALMEIDA, C.W.D. et al. **Web System for Digital Mammograms Analysis**. Proceedings of the Conference on Medical Cybernetics in Clinical Practice in Burdenko Main Military Clinical Hospital, Vol. 2, pp.110-119, Moscow, Russia, 2004b.

ALMEIDA, C.W.D. et al. **Web System for Medical Support using Virtual Reality**. Proceedings of the Second International Conference Biomedical Engineering, Innsbruck, Austria, 2003a.

ALMEIDA, S. M. C., BRASIL, L. M., CARVALHO, H. S., FERNEDA, E., SILVA, R. P. **Proposal of a Case-Based Reasoning Model for IACVIRTUAL Project**. ISBN: 972-8924-13-5. Proceedings of the IADIS MCCSIS 2006 - International Association for Development of the Information Society (Multi Conference on Computer Science and Information Systems / Topics of Health Informatics), p. 312-314, May 15-19, 2006a.

ALMEIDA, S. M. C., BRASIL, L. M., CARVALHO, H. S., FERNEDA, E., SILVA, R. P. **The Diagnosis Support System for Ischemic Cardiopathy: A Case Study in the Context of IACVIRTUAL Project**. ISSN: 1727-1983. IFMBE (International Federation for Medical and Biological Engineering) Proceedings. Volume 14, 2006 - World Congress on Medical Physics and Biomedical Engineering. p. 3541-3544, Korea, Seoul, August 27-September 01, 2006b.

ALVES, A. C. **Algoritmos Genéticos Aplicados ao Planejamento da Distribuição de Energia Elétrica em Curitiba e Região Metropolitana - A Localização das Subestações**. Dissertação de Mestrado do programa de Pós-graduação em Métodos Numéricos em Engenharia, Departamento de Matemática, Universidade Federal do Paraná, Curitiba, 2002.

ARQUIVOS BRASILEIROS DE CARDIOLOGIA. **Diretrizes de Doença Coronariana Crônica Angina Estável**. Vol. 83, Suplemento I, Setembro 2004.

AUSTIN, S. **An Introduction to Genetic Algorithms**. AI Expert, vol. 1, No. 1, pp. 48-53, 1990.

AZEVEDO, F. M. **Algoritmos Genéticos em Redes Neurais Artificiais**. V Escola de Redes Neurais - ITA - Conselho Nacional de Redes Neurais - pp. c091-c121, São José dos Campos - SP, 19 de julho de 1999.

AZEVEDO, F.M., BRASIL, L.M., OLIVEIRA, R.C.L. **Redes Neurais com aplicação em controles e em sistemas especialistas**. Bookstores, Florianópolis, SC, 2000.

BARRETO, J.M. **Conexionismo e a Resolução de Problemas**. Trabalho apresentado no Concurso Público de Prof. Titular - Departamento de Informática e Estatística - Centro Técnico e Científico - UFSC, 1996.

BEDDOE, G., PETROVIC, S. **A novel approach to finding feasible solutions to personnel rostering problems**. Proceedings of the 14th Annual Conference of the Production and Operation Management Society, Savannah, Georgia, United States, 2003.

BEDDOE, G., PETROVIC, S. **Selecting and weighting features using a genetic algorithm in a case-based reasoning approach to personnel rostering**. Accepted for publication in the European Journal of Operational Research, 2004.

BRAGA JR., M.S. **Proposta de Modelo RBC para a Recuperação Inteligente de Jurisprudência na Justiça Federal**. Dissertação de Mestrado em Engenharia de Produção, Universidade Federal de Santa Catarina, 2001.

CAMPOS, M.M, SAITO, K. **Sistemas Inteligentes em Controle e Automação de Processos**. Editora Ciência Moderna, Brasil, 2004.

CORRÊA, E.S. **Algoritmos Genéticos e Busca Tabu Aplicados ao Problema das P-medianas**. Dissertação de Mestrado, Programa de Pós-graduação em Métodos Numéricos em Engenharia, UFPR, Curitiba, 2000.

COSTA, E.B., LOPES, M.A., FERNEDA, E. **MATHEMA: A Learning Environment Based on a Multi-Agent Architecture**. In: J. Wainer and A. Carvalho (Eds.), Proceedings of 12th Brazilian Symposium on Artificial Intelligence, Lecture Notes in Artificial Intelligence, Vol. 991, pp. 141-150. 1995.

DAVENPORT, T.H., PRUSAK, L. **Conhecimento Empresarial: Como as Organizações Gerenciam o seu Capital Intelectual**. Editora Campus, Rio de Janeiro, 1998.

DAVIS, L. **Handbook of Genetic Algorithms**. New York, Van Nostrand Reinhold, 1991

DEMIROZ, G., GUVENIR, H.A. **Genetic Algorithms to Learn Feature Weights for the Nearest Neighbor Algorithm**. Proceedings of the Belgian--Dutch Conference on Machine Learning (BENELEARN--96), p.117--126, 1996.

DRUMMOND, I.N. **Uso de Relações de Similaridade para Tratamento de Consistência e Cobertura em Sistemas de Regras Difusas**. Tese de Mestrado em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, INPE, Brasil, 2003.

ESHELMAN, L.J., SCHAFFER, J.D. **Real-coded genetic algorithms and interval schemata**. In Whitley, L. D., editor, Foundations of Genetic Algorithms 2, pages 187--202. Morgan-Kaufmann, San Mateo, CA, 1993.

GARIS, H. **Exploring GenNet Behaviors - Using Genetic Programming to Explore Qualitatively New Behaviors in Recurrent Neural Networks**. In: IJCNN - International Joint Conference on Neural Networks, Baltimore, MD, USA, Vol. 3 p.547-552, 1992.

GOLDBERG, D.E. **Genetic Algorithms in Search, Optimization and Machine Learning**. Addison Wesley, CA, 1989.

GRFENSTETTE, J.J. **Optimization of Control Parameters for Genetic Algorithms**. IEEE Transactions on Systems, Man and Cybernetics, v. 16, n. 1, p. 122-128, 1986.

HALL, M.A. **Correlation-based Feature Selection for Machine Learning**. Tese de Doutorado do Departamento de Ciência da Computação da Universidade de Waikato, Hamilton, Nova Zelândia.

HAUPT, R.L., HAUPT, S.E. **Practical Genetic Algorithms**. New York: Wiley Interscience, 1998.

HOLLAND, J. **Adaptation in Natural and Artificial Systems**. University of Michigan Press, 1975.

HOLLAND, J. **Outline for a Logical Theory of Adaptative Systems**. In: Journal of the Association for Computing Machinery, 3, p.297-314, 1962.

HUNT, J.E., FELLOWS, A. **Introducing an Immune Response into a CBR system for Data Mining**. BCS ESG'96 Conference, 1996.

KASSIRER, J.P., GORRY G.A. **Clinical Problem Solving: a Behavioral Analysis**. Annals of Internal Medicine, Vol. 89, n. 1, p. 245-255, 1978.

KOLODNER, J. **Case-Based Reasoning**. Morgan Kaufmann, San Francisco, California, 1993.

LACERDA, E.G., CARVALHO, A.C.P.L.F. **Introdução aos Algoritmos Genéticos**. In: Anais do XIX Congresso Nacional da Sociedade Brasileira de Computação, Vol. 2 pp. 51-126, 1999.

LAROSE, D.T. **Discovering Knowledge in Data: An Introduction to Data Mining**. John Wiley and Sons, Hoboken, NJ, 2005.

LOPES, H. S. **Analogia e Aprendizado Evolucionário: Aplicação em Diagnóstico Clínico**. Tese de Doutorado em Engenharia Elétrica, Universidade Federal de Santa Catarina, 1996.

MARTINS, A. **Computação baseada em Casos: Contribuições Metodológicas aos Modelos de Indexação, Avaliação, Ranking, e Similaridade de Casos**. Tese de Doutorado em Engenharia Elétrica, Universidade Federal da Paraíba, 2000.

MENDES, J.R.P. **Raciocínio Baseado em Casos Aplicado ao Projeto de Poços de Petróleo**. Tese de Doutorado em Engenharia Mecânica, Universidade Estadual de Campinas, Unicamp, 2001.

MICHALEWICZ, Z. **Genetic Algorithms + Data Structures = Evolution Programs**. Berlin: Springer-Verlag, 1992.

MITCHEL, T.M. **Machine Learning**. McGraw-Hill, USA, 1997.



- MORESI, E.A.D. **Gestão da Informação e do Conhecimento**. In: Tarapanoff, K. (Ed): *Inteligência Organizacional e Competitiva*, Universidade de Brasília, 2001.
- MORESI, E.A.D. **Manual de Metodologia da Pesquisa**. Universidade Católica de Brasília, Brasília-DF, 2003.
- NONAKA, I.M, TAKEUCHI, H. **Criação de Conhecimento na Empresa**. Rio de Janeiro, Ed. Campus, 1997.
- PORTAL MÉDICO. Desenvolvido por Conselho Federal de Medicina. Disponível em: <http://www.cremeb.cfm.org.br>. Acesso em: 18 de abril de 2005.
- QUINLAN, J.R. **An Empirical Comparison of Genetic and Decision-tree Classifiers**. In: *Proceedings of the Fifth International Machine Learning Conference*, San Mateo, CA, p. 135-141, 1988.
- QUINLAN, J.R. **C4.5: Programs for Machine Learning**. Morgan Kaufmann, San Mateo, CA, 1993.
- RABELO JR., A. et al. **Um Sistema Especialista para Diagnóstico de Cardiopatias Isquêmicas**. *Revista Informédica*, vol 1(1): 5-11, 1993.
- RADCLIFFE, N.J. **Formal analysis and random respectful recombination**. In: R. K. Belew and L. B. Booker (Eds.), *Proceedings of the Fourth International Conference on Genetic Algorithms* (pp. 222--229), 1991.
- RODRIGUES, G.J.O. **Ferramentas Computacionais para Otimização e Síntese de Sistemas Híbridos de Risers Baseados no Conceito de Bóia de Superfície**. Tese de Doutorado em Engenharia Civil, Universidade Federal do Rio de Janeiro, 2004.
- ROJAS J.C.C. **Sistema Especialista Híbrido de Apoio à Decisão de uma Equipe Clínico-Cirúrgica na Definição de Conduta Terapêutica em Pacientes Coronariopatas**. Dissertação de Mestrado em Engenharia Biomédica, Universidade Federal da Paraíba, 2003.
- SCHANK, R.C. **Dynamic Memory: A theory of learning in computers e people**. New York, Cambridge University Press, 1982.
- SILVA, A.J.M. **Implementação de um Algoritmo Genético Utilizando o Modelo de Ilhas**. Dissertação de Mestrado do programa de pós-graduação em Engenharia da UFRJ - Universidade Federal do Rio de Janeiro, 2005b.
- SILVA, J.D.S. **Uma Abordagem Híbrida por Dempster-Shafer e Algoritmos Genéticos para o Problema de Correspondência em Estereoscopia**. Tese de Doutorado em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, 1999.
- SILVA, L.N.C. **Engenharia Imunológica: Desenvolvimento e Aplicação de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais**. Tese de Doutorado em Engenharia Elétrica, Unicamp, Campinas, SP, 2001.
- SILVA, R.P. **Modelo de Apoio ao Diagnóstico no Domínio Médico Aplicando Raciocínio Baseado em Casos**. Dissertação de Mestrado em Gestão do Conhecimento e da Tecnologia da Informação. Universidade Católica de Brasília, 2005a.

SILVA, R.P., FERNEDA, E., BRASIL, L.M., SILVA, A.P.B., ROJAS, J.C.C., MELO, C.B., POLITO, R.F. **Aplicação de Raciocínio Baseado em Casos num Sistema de Apoio à Educação e Diagnóstico Médico**. In: III Congresso Latino-Americano de Engenharia Biomédica (CLAEB2004), João Pessoa, Paraíba, Brasil, 2004.

SRINIVAS, M., PATNAIK, L.M. **Adaptive probabilities of crossover and mutation in genetic algorithm**. IEEE Transactions on Systems, Man, and Cybernetics, vol.24, no.4, pp.656-667, 1994.

STOLLENWERK, M.F.L. **Gestão do Conhecimento: conceitos e modelos**. In: Tarapanoff, K. Inteligência Organizacional e Competitiva. Universidade de Brasília, 2001.

TANOMARU, J. **Motivação, Fundamentos e Aplicações de Algoritmos Genéticos**. In: Anais do 2º Congresso Brasileiro de Redes Neurais - CBRN2005, Curitiba, PR, Brasil, [s.p.], 1995.

TEIXEIRA FILHO, J. **Gerenciando conhecimento**. Rio de Janeiro, Ed. SENAC, 2001.

THÉ, M.A.L. **Raciocínio Baseado em Casos. Uma Abordagem Fuzzy para Diagnóstico Nutricional**. Dissertação de Mestrado em Engenharia de Produção, Universidade Federal de Santa Catarina, 2001.

TSUNODA, D.F. **Abordagens Evolucionárias para a Descoberta de Padrões e Classificação de Proteínas**. Tese de Doutorado em Engenharia Biomédica, Centro Federal de Educação Tecnológica do Paraná, 2004.

VERGARA, S.C. **Projetos e relatórios de pesquisa em administração**. São Paulo: Atlas, 3ª ed., 2000.

WANG, Y., ISHII, N. **A genetic algorithm for learning weights in a similarity function**. In George D. Smith and Nigel C. Steele, editors, Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms, pages 206-209, Norwich, UK, 2.-4. Springer-Verlag, Berlin, 1997.

WANGENHEIM, C.H., WANGENHEIM, A.V. **Raciocínio Baseado em Casos**. São Paulo, Ed. Manole, 2003.

WETTSCHERECK, D., AHA, D. **Weighting features**. In: Proceeding of the 1st International Conference on Case-Based Reasoning. Springer, 1995.