Proceedings of the V Brazilian Conference on Neural Networks - V Congresso Brasileiro de Redes Neurais
pp. 241-246, April 2-5, 2001 - Rio de Janeiro - RJ - Brazil

Mineração de regras *Fuzzy* em bancos de dados por Algoritmo Genético

Jaime de Britto, Heitor S. Lopes
CPGEI/CEFET-PR
Av. Sete de Setembro, 3165 80.230-901 Curitiba - PR
britto@cpgei.cefetpr.br hslopes@cpgei.cefetpr.br

Abstract

Diagnostic support systems are a clear example of applications where knowledge involved is mostly heuristic. Acquiring that knowledge from historical databases is a challenge approached by some genetic algorithms' methods. Moreover, most biomedical databases are composed of non deterministic data, requiring the implementation of a system capable of representing and manipulating these information – these are the fuzzy systems. The present work aims joining fuzzy and genetic algorithms' techniques in a suitable structure for medical diagnostic support, without getting bounded to the respective database. To evaluate this development, results obtained on Wisconsin malignancy of breast cancer diagnostic and on kidney disorders databases are shown.

1. Introdução

Diversos métodos e técnicas de otimização têm sido utilizados com sucesso em sistemas com comportamento bem definidos. Este sucesso motivou a aplicação direta das mesmas abordagens para a modelagem e análise de sistemas cujas decisões principais são dadas por julgamentos heurísticos. Infelizmente, o progresso foi muito menor que o esperado. A causa disto possivelmente está ligada às seguintes questões: Pode-se encontrar uma boa solução rapidamente, satisfazendo um bom nível de desempenho? Trabalhando com dados imprecisos, é possível projetar um algoritmo eficiente para obter uma boa solução? Como atingir uma solução aceitável baseada em mecanismos de raciocínio humano? O que se conhece do sistema é o suficiente para se obter uma boa solução? Estas questões são difíceis de se responder. Para sistemas complexos e de comportamento não linear, muitas vezes é mais importante buscar uma boa solução rapidamente do que buscar a solução ótima. Quando se busca melhorar o desempenho de um determinado processo, procurando-se atingir um ótimo global, um paradigma computacional bastante utilizado é o algoritmo genético. Algoritmos genéticos (AG) são métodos de busca que têm ganho grande atenção devido ao seu potencial como técnica de otimização. Eles são algoritmos de busca, que não necessariamente encontram uma solução ótima para qualquer problema, mas podem encontrar boas soluções para problemas que são difíceis de resolver por qualquer outra técnica conhecida. Um bom exemplo disto, é o aprendizado de máquina para sistemas especialistas, tais como sistemas de apoio ao diagnóstico. Uma das tarefas mais importantes e difíceis, neste caso, é a extração do conhecimento do especialista. A dificuldade encontrada na extração deste conhecimento na forma de regras tem levado diversos pesquisadores a desenvolver técnicas automatizadas para extração do conhecimento necessário ao sistema. A técnica explorada neste trabalho é a aplicação de um algoritmo genético para treinamento de um sistema *fuzzy* de apoio ao diagnóstico médico. A lógica *fuzzy* é um paradigma computacional que provê uma ferramenta matemática para representação e manipulação de informações, que se parece com os processos de comunicação e raciocínio humanos[1].

Em um processo de aprendizado de máquina para a geração de um sistema *fuzzy*, pode-se distinguir os seguintes componentes:

1. Um método de geração de regras *fuzzy* capazes de incluir o conhecimento subjacente a um conjunto de exemplos.
2. Um método de simplificação:
   2.1. Detectando e eliminando regras que se contradizem.
   2.2. Encontrando um conjunto final de regras *fuzzy* capazes de aproximar o comportamento de entrada/saída do sistema real.
3. Um método de sintonia do conjunto de regras obtido no passo 2.

1.1. Objetivo

Este trabalho descreve uma abordagem de treinamento de um sistema *fuzzy* de apoio ao diagnóstico médico, através de um algoritmo evolucionário baseado em exemplos e compara os resultados obtidos com os disponíveis na literatura.

A maioria dos métodos de treinamento propostos é dependente dos dados com os quais o sistema foi treinado, descobrindo sistemas com desempenho próximo a 100% de acerto. Nesta ótica, ao se aplicar a mesma técnica em outra base de casos, onde por exemplo, as classes não aparecem com a mesma freqüência, o desempenho do sistema descoberto tende a ser insatisfatório.

A técnica de treinamento, desenvolvida neste trabalho pode ser aplicada a qualquer base de casos com atributos contínuos, para construir sistemas *fuzzy*. Podendo ser aplicada inclusive em bases de casos onde as classes não aparecem nas mesmas proporções.

Em linhas gerais, pode-se descrever o sistema da seguinte forma: (1) o algoritmo genético gera um conjunto de regras sintonizadas e simplificadas e as funções de pertinência; (2) o sistema *fuzzy* avalia estas regras em uma partição de uma base de dados e retorna o resultado desta avaliação para o algoritmo genético como sendo o *fitness* daquele conjunto de regras gerados; (3) O algoritmo genético procura pelo melhor conjunto de regras tentando maximizar o *fitness* retornado pelo sistema fuzzy; (4) ao final da execução do algoritmo genético, avalia-se o melhor conjunto de regras gerado em uma outra partição da base de dados.

Para avaliar a técnica aplicada, foram utilizadas duas bases de dados da área médica. A primeira - clássica na literatura - é a base de casos de diagnóstico da malignidade de câncer de mama de Wisconsin. A segunda - menos utilizada - é a base de casos de desordens do fígado do *BUPA Medical Research Ltd.*

1.2. A base de casos de malignidade do câncer de mama

Este banco de dados contém informações relevantes nos exames citológicos realizados em biópsias de tumores de mama. Nove atributos são descritos:

1. Espessura do conjunto de células;
2. Uniformidade de tamanho das células;
3. Uniformidade de forma das células;
4. Adesão marginal;
5. Tamanho da célula epitelial simples;
6. Núcleos descobertos;
7. Cromatina Branda;
8. Nucléolos normais;
9. Mitoses.

Para cada um destes itens, é atribuído um grau de anormalidade que varia de 1 a 10, o grau 1 representa 100% de normalidade, o grau 10 representa o maior grau de anormalidade possível.

Após a análise, o patologista faz a conjunção destes valores para diagnosticar o tipo de câncer encontrado, se benigno ou maligno[2].

Os diagnósticos na base de dados WBCD (*Wisconsin Breast Câncer Database)* foram fornecidos por especialistas na área. A base de dados consiste de 683 casos, com cada entrada representando a classificação para um certo conjunto de valores.

Existem diversos estudos baseados nesta base de dados. [3] usaram técnicas de Redes Neurais, obtendo uma taxa de classificação de 99,6% quando esta base de dados continha 487 casos. Entretanto sua solução apresenta grande complexidade, por exemplo, as decisões de diagnóstico são essencialmente caixas-pretas, sem nenhuma explicação de como se chegou àquela conclusão. Com a interpretabilidade em mente como o objetivo principal, um número de pesquisadores tem aplicado o método de extração de regras tradicionais de Redes Neurais[4,5,6]. Seus resultados são animadores, exibindo bom desempenho e reduzido número de regras e variáveis de entrada relevantes. Entretanto, estes sistemas de regras tradicionais não são capazes de fornecer ao usuário uma medida de confiança para a decisão tomada. O trabalho proposto por [7] mostrou que é possível obter alto desempenho com interpretabilidade e medida de confiança.

1.3. A base de casos de desordens do fígado

Esta base de casos é composta dos seguintes atributos:

1. Volume corpuscular médio;
2. Fosfatase alcalina;
3. Alanina aminotransferase (SGPT);
4. Aspartato aminotransferase (SGOT);
5. Gama-glutamil transpeptidase;
6. Número de doses de bebidas alcoólicas tomadas por dia

As primeiras cinco variáveis são todas testes de sangue, as quais, acredita-se, serem sensitivas a desordens do fígado causadas por consumo excessivo de álcool.

A base de exemplos de desordens do fígado apresenta ao todo 345 casos, dos quais 145 são da classe sadio e 200 são da classe doente.

O único trabalho conhecido sobre esta base de casos é o citado em *PC/BEAGLE User's Guide* [8].

2. Metodologia

2.1. Estrutura do sistema *fuzzy*

Quando da definição deste trabalho, algumas premissas básicas foram consideradas para direcionar o desenvolvimento:

- Pequeno número de regras. O número de regras de um sistema *fuzzy* é dado por $(Nv)^n$, onde Nv é o número de variáveis e n é o número de predicados. Isto leva a uma explosão combinatória à medida que cresce o número de variáveis. Entretanto, sistemas com menos que cinco regras têm demonstrado obter alto desempenho [4,7];
- Pequeno número de variáveis. Regras com menos que cinco antecedentes, em média, são adequadas e trazem compreensibilidade à solução[4,6,7];

2.1.1. Regras e parâmetros *fuzzy*

Na geração, simplificação e sintonia da base de regras e de dados *fuzzy* foi utilizado, com algumas alterações, o AG proposto por [9], conforme descrito na subseção 2.2.

2.1.2. Método de *defuzzificação*:

O método de *defuzzificação* utilizado foi a Média dos Máximos:

$$v_0 = \sum_{j=1}^{J} \frac{v_j}{J} \quad (1)$$

Onde:

- v_j → valor do suporte para o qual a função de pertinência atinge o máximo;
- J → total de valores para os quais a função atinge o máximo.

2.1.3. Operações utilizadas:

No sistema *fuzzy* foram utilizados a função de implicação de Lukasiewicz [10] (equação 2) e o operador de conjunção T MIN (equação 3): N primeira, a pertinência da saída da regra será o mínimo entre 1 e o complemento da pertinência dos antecedentes mais a pertinência do conseqüente; No operador de conjunção, a pertinência dos antecedentes será a menor pertinência entre eles.

$$\mu_R(u,v) = min(1,1-\mu_A(u)+\mu_B(u)) \quad (2)$$

$$\mu A \cap B = \mu A(u) t \mu B(u) \leq min(\mu A(u), \mu B(u)) \quad (3)$$

2.2. O Algoritmo Genético

2.2.1. Codificação

O algoritmo genético busca quatro tipos de parâmetros: atributos relevantes, valores de função de pertinência de entrada, antecedentes das regras e o conseqüente da regra; e é construído conforme segue:

- Parâmetros da função de pertinência. Existem *j* variáveis (v_1...v_j), cada uma com dois parâmetros *P* e *d*, definindo, respectivamente, o ponto de início e a largura da região de interseção das funções de pertinência (figura 1)
- Antecedentes. A i-ésima regra tem a forma:
- **SE** (v_1 é A_1^i) **E ... E** (v_j é A_j^i) **ENTÃO** (saída é D_k), onde A_j^i representa a função de pertinência aplicável à variável v_j e D_k representa o diagnóstico do caso. A_j^i pode ter os valores: 1 (BAIXO), 2 (ALTO), 0 ou 3 (IRRELEVANTE). D_k pode assumir qualquer uma das *k* classes.
- Variáveis relevantes são procuradas implicitamente quando o algoritmo escolhe o valor *IRRELEVANTE* para determinada variável; neste caso a respectiva variável é considerada irrelevante. Por exemplo, a regra:
  SE (v*1* é ALTO) **E** (v*2* é IRRELEVANTE) **E** (v*3* é IRRELEVANTE) **E** (v*4* é BAIXO) **E** (v*5* é IRRELEVANTE) **E** (v*6* é IRRELEVANTE) **E** (v*7* é IRRELEVANTE) **E** (v*8* é BAIXO) **E** (v*9* é IRRELEVANTE) **ENTÃO** (saída é D_1), é interpretada como:
  SE (v*1* é ALTO) **E** (v*4* é BAIXO) **E** (v*8* é BAIXO) **ENTÃO** (saída é D_2).

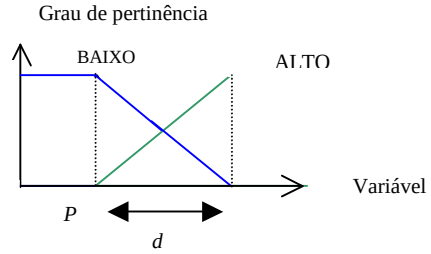

Figura 1: Exemplo de uma variável fuzzy com dois possíveis valores ALTO e BAIXO

A tabela 1 demonstra os parâmetros de codificação de um caso contendo: o início das funções de pertinência de 9 atributos previsores *(P)* com valores variando de 1 a 10; a largura de interseção das funções de pertinência *(d)* de cada atributo previsor, variando de 1 a 5; um conjunto de regras *(Cr)* contendo 5 regras, cada uma codificando um termo lingüístico para cada um dos 9 atributos previsores – 1 (BAIXO), 2 (ALTO), 0 ou 3 (IRRELEVANTE); e o conseqüente de cada regra *(C)*. Para codificarmos o início das funções de pertinência de um atributo previsor, precisamos de 4 bits, a largura de interseção das funções de pertinência necessita de 3 bits, cada termo lingüístico utiliza 2 bits, o conseqüente das regras, como as bases de dados utilizadas apresentam apenas duas classes, necessita de apenas 1 bit. Tais parâmetros, juntos, formam um único indivíduo. O tamanho total do cromossomo, em bits, é:

*9 * 4 + 9 * 3 + 9 * 5 * 2 + 1 *5.*

Assim, o espaço de busca para um cromossomo é:

$2^{158} = 3{,}65 \times 10^{47}$

Tabela 1: Parâmetros codificando um indivíduo.

Parâmetros	Valores	Bits	Quantidade de Parâmetros	Nº total de bits
P	[1...10]	4	9	36
D	[1...5]	3	9	27
Cr	[0...3]	2	9 * 5	90
C	[0,1]	1	*1 * 5*	*5*

A figura 2 demonstra um exemplo de cromossomo codificando duas regras com as funções de pertinência de seus atributos.

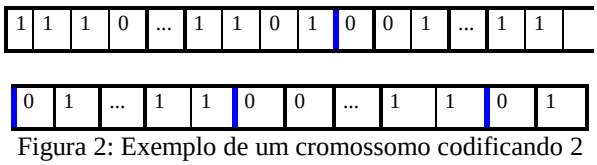

Figura 2: Exemplo de um cromossomo codificando 2 regras, com as funções de pertinência de seus atributos.

Obs:

- Os valores de início e largura das funções de pertinência foram separados das regras porque são comuns a todas as regras.

2.2.2. Função de *Fitness*

A função de *fitness* proposta por [9] combina três critérios: (1) Fc: Desempenho de classificação, computada como a porcentagem de casos corretamente diagnosticados; (2) Fe: a diferença quadrática (equação 4) entre o valor contínuo de avaliação (no intervalo {2 – 4}) e o diagnóstico discreto correto dado pela base casos; e (3) Fv: o número médio de variáveis por regra. A função de *fitness* é dada por $F = Fc - \alpha Fv - \beta Fe$, onde α e β são valores derivados empiricamente.

$$\sqrt[2]{(f-a)^2} \qquad (4)$$

Onde *f* representa o valor contínuo de avaliação e *a* representa o diagnóstico discreto correto dado pela base de casos.

Porém, [11] demonstra que esta abordagem apresenta um desempenho insatisfatório quando aplicada a casos que apresentem distribuição de classes desproporcional e propõem o uso de dois índices bastante usuais na área médica, para quantificar o desempenho de um teste classificador: a sensibilidade (S) e a especificidade (E) mostrados nas equações 5 e 6. A sensibilidade identifica a proporção dos casos verdadeiramente positivos classificados como tal pelo teste, enquanto que a especificidade representa a proporção dos casos verdadeiramente negativos que o teste identifica como negativos.

$$S = \frac{vp}{vp + fn} \qquad (5)$$

$$E = \frac{vn}{vn + fp} \qquad (6)$$

Onde vp, vn, fn, fp, significam, respectivamente, verdadeiro-positivo, verdadeiro-negativo, falso-positivo, falso-negativo. Os dois primeiros índices representam os acertos enquanto os dois últimos representam os erros.

Substituiu-se então, a função de *fitness* pela proposta, que é a seguinte:

$$Fitness = S \times E \qquad (7)$$

2.2.3. Operadores e outros parâmetros

Para desenvolver o sistema de inferência *fuzzy*, foi usado um algoritmo genético com uma população fixa de 100 indivíduos e método de seleção do tipo torneio estocástico - com o tamanho do torneio sendo 5. O algoritmo termina quando atinge o número máximo de gerações (250). As probabilidades de aplicação de *crossover* e mutação são 0.5 e 0.01 respectivamente. Além disto, ainda foram usados redução de incesto e fator de *crowding* para o controle da pressão seletiva, com valores 1 e 2, respectivamente.

A definição do tamanho e do valor máximo de cada alelo é feita automaticamente, a partir de um arquivo de definições da base de exemplos que será utilizada.

3. Resultados

3.1. Estatísticas

A avaliação de desempenho foi feita de duas formas. Primeiro, de acordo com a partição dos dados em dois subconjuntos distintos: conjunto de treinamento, contendo 75% dos casos e conjunto de teste (ou avaliação), contendo 25% dos casos, ambos selecionados aleatoriamente, conforme descritos nas tabelas 2 e 3.

Tabela 2: Partição dos dados da base de dados de câncer de mama.

Partição	Benigno	Maligno	Total
Completa	444	239	683
Treinamento	312	200	512
Teste	133	38	171

Tabela 3: Partição dos dados da base de dados de desordens do fígado.

Partição	Benigno	Maligno	Total
Completa	145	200	345
Treinamento	112	146	258
Teste	33	54	87

A segunda forma de avaliação utilizou a técnica de validação cruzada [12], com os dados separados aleatoriamente em 5 subconjuntos, respeitando a proporcionalidade das classes.

A técnica de validação cruzada divide o conjunto todo de casos em *k* subconjuntos, seleciona *k*-1 subconjuntos para treinar o sistema e computa a taxa de acerto do conhecimento descoberto no subconjunto que ficou de fora. Em seguida, seleciona um subconjunto que ainda não foi utilizado para teste e novamente treina o sistema nos *k*-1 subconjuntos restantes, computando a taxa de acerto no subconjunto selecionado. Estes passos se repetem até que todos os *k* subconjuntos sejam utilizados para teste e os outros *k*-1 subconjuntos para treinamento. Ao final, computa-se a taxa de acerto do algoritmo como a média aritmética das taxas de acerto das *k* iterações.

O número de regras foi fixado para ser entre duas e cinco regras.

A tabela 4 apresenta o desempenho obtido pelo AG em 100 execuções sobre a base de casos de câncer de mama, com os dados divididos em dois subconjuntos, enquanto que a tabela 5 mostra os resultados utilizando-se a técnica de validação cruzada.

A tabela 6 compara o melhor sistema descoberto com os sistemas obtidos por quatro outras abordagens de diagnóstico baseado em regras sobre a base de casos de câncer de mama. As primeiras três abordagens – de [4], [5] e [6] – envolvem bases de regras tradicionais extraídas de Redes Neurais; a última abordagem é o trabalho de [9].

A tabela 7 compara os resultados obtidos, aplicando-se a técnica proposta por [9] e a aplicada neste trabalho sobre a base de casos de desordens do fígado.

As figuras 3 e 4 mostram os parâmetros do sistema *fuzzy* e as regras do melhor conjunto descoberto, aplicado à base casos de câncer de mama, utilizando-se respectivamente, a partição em dois subconjuntos e a validação cruzada.

As figuras 5 e 6 mostram os parâmetros do sistema fuzzy e as regras do melhor conjunto descoberto, aplicado à base casos de desordens do fígado, utilizando-se respectivamente, a partição em dois subconjuntos e a validação cruzada.

Tabela 4: Desempenho do sistema nas fases de treinamento e teste, utilizando partição dos dados em dois subconjuntos (75% e 25%) para a base de dados de câncer de mama.

Treinamento	Teste	Nº regras
93,76%	98,26%	2
94,07%	98,84%	3
93,59%	98,84%	4
94,07%	96,51%	5

Tabela 5: Desempenho do sistema nas fases de treinamento e teste, utilizando a validação cruzada para a base de dados de câncer de mama.

Desempenho		
Treinamento	Teste	Nº regras
98,21%	97,78%	2
95,84%	95,64%	3
94,94%	94,77%	4
95,98%	95,64%	5

Tabela 6: Comparativo de desempenho entre os vários trabalhos existentes na literatura e o trabalho desenvolvido.

Nº Regras	Setiono	Setiono e Liu	Taha e Ghosh	Penã-Reyes	Este Trabalho	
					75/25%	Valid. cruzada
1	95.14%	--	--	97.07%	--	--
2	--	--	--	97.37%	98,26%	97,78%
3	97.14%	97.21%	--	97.80%	98,84%	95,64%
4	--	--	--	97.80%	98,84%	94,77%
5	--	--	96.19%	97.51%	96,51%	95,64%

Tabela 7: Comparativo de desempenho da técnica proposta por [9] e a aplicada neste trabalho, para a base de dados de desordens do fígado.

Nº de regras	Este Trabalho				Peña-Reyes	
	75/25%		Valid. Cruzada		75/25%	
	Trein. (%)	Teste (%)	Trein. (%)	Teste (%)	Trein. (%)	Teste (%)
4	55,97	72.54	53,26	71,52	62,05	62,14

Parâmetros									
	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	7	2	8	2	1	3	7	4	2
d	1	3	2	5	1	3	3	5	3

Regras

R1: **SE** v_1 = BAIXO **E** v_2 = BAIXO **E** v_5 = ALTO **E** v_8 = BAIXO **ENTÃO** diagnóstico = Benigno

R2: **SE** v_1 = BAIXO **E** v_5 = ALTO **E** v_6 = ALTO **E** v_9 = BAIXO **ENTÃO** diagnóstico = Maligno

R3: **SE** v_1 = BAIXO **E** v_4 = ALTO **E** v_5 = ALTO **E** v_6 = BAIXO **ENTÃO** diagnóstico = Maligno

Figura 3: base de dados e de regras do melhor sistema descoberto, aplicado à base casos de câncer de mama, utilizando-se a partição da base de casos em dois subconjuntos.

Parâmetros									
	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	2	3	0	7	2	6	2	9	7
d	5	3	4	4	1	1	2	1	1

Regras

R1: **SE** v_1 = BAIXO **E** v_3 = ALTO **E** v_4 = BAIXO **E** v_6 = BAIXO **E** v_8 = BAIXO **ENTÃO** diagnóstico = Benigno

R2: **SE** v_2 = ALTO **ENTÃO** diagnóstico = Maligno

Figura 4: base de dados e de regras do melhor sistema descoberto, aplicado à base casos de câncer de mama, utilizando-se a validação cruzada.

Parâmetros						
	v_1	v_2	v_3	v_4	v_5	v_6
P	20	53	13	10	17	3
d	91	59	100	44	53	10

Base de Regras

R1: **SE** v_4 = BAIXO **E** v_5 = ALTO **E** v_6 = BAIXO **ENTÃO** diagnóstico = Doente

R2: **SE** v_1 = BAIXO **E** v_2 = ALTO **E** v_3 = ALTO **E** v_4 = ALTO **E** v_5 = ALTO **E** v_6 = BAIXO **ENTÃO** diagnóstico = Doente

R3: **SE** v_2 = ALTO **E** v_5 = ALTO **ENTÃO** diagnóstico = Doente

R4: **SE** v_1 = ALTO **E** v_2 = ALTO **E** v_3 = ALTO **E** v_4 = BAIXO **ENTÃO** diagnóstico = Sadio

Figura 5: base de dados e de regras do melhor sistema descoberto, aplicado à base casos de desordens do fígado, utilizando-se a partição da base de casos em dois subconjuntos.

Parâmetros						
	V_1	v_2	v_3	v_4	v_5	V_6
P	93	14	19	8	17	5
D	7	107	78	37	10	3

Regras

R1: **SE** v_1 = BAIXO **E** v_4 = ALTO **ENTÃO** diagnóstico = Doente

R2: **SE** v_4 = ALTO **E** v_6 = BAIXO **ENTÃO** diagnóstico = Doente

R3: **SE** v_2 = ALTO **E** v_3 = BAIXO **E** v_4 = BAIXO **E** v_5 = ALTO **ENTÃO** diagnóstico = Doente

R4: **SE** v_1 = BAIXO **E** v_2 = ALTO **E** v_3 = ALTO **E** v_4 = ALTO **ENTÃO** diagnóstico = Sadio

Figura 6: base de dados e regras do melhor sistema descoberto, aplicado à base casos de desordens do fígado, utilizando-se a validação cruzada.

4. Conclusões

A simples observação dos valores atingidos nos comparativos permite concluir que: (1) a técnica de se computar apenas a taxa de acerto, baseando-se unicamente nos casos bem diagnosticados, não é robusta e, apesar de apresentar alto desempenho em bases de casos onde as classes aparecem com a mesma freqüência, não apresenta poder de classificação satisfatório quando as classes não aparecem com a mesma freqüência, ao se aplicar a sensibilidade e a especificidade como forma de avaliação de acerto, atingiu-se melhores resultados; (2) os níveis de desempenho atingidos com a divisão dos dados em duas partições (75% treinamento / 25% teste) não mostraram-se confiáveis, pois houve variação dos resultados conforme a partição selecionada, como estamos lidando com um algoritmo de otimização, este pode vir a selecionar uma partição que tem quase todos os seus casos cobertos pelas regras, mas que não necessariamente reflete a realidade da base de dados como um todo, ao se aplicar a técnica de validação cruzada, o sistema é praticamente treinado e testado com todos os dados.

O sistema *fuzzy* apresentou um desempenho similar aos anteriormente desenvolvidos, com a vantagem de que suas regras possuem um maior grau de interpretabilidade, oferecendo ao usuário mais confiança nas decisões tomadas.

O algoritmo genético mostrou-se ser uma poderosa ferramenta de otimização, de fácil implementação, atingindo boas soluções rapidamente.

Referências

[1] R. R. Yager and L. A. Zadeh. *Fuzzy Sets, Neural Networks, and Soft Computing*. Van Nostrand Reinhold, New York, 1994.

[2] W. N. Street, W. H. Wolberg and O. L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. *International Symposium on Electronic Imaging: Science and Technology*, volume 1905, pages. 861-870, San Jose, California, 1993.

[3] K. P. Bennett and O. L. Mangasarian. Neural Network training via linear programming. *Advances in Optimization and Parallel Computing*, Elsevier Science, pages 56-57, 1992.

[4] R. Setiono. Extracting rules from pruned neural networks for breast cancer diagnosis. *Artificial Intelligence in Medicine*, pages. 37-51, 1996.

[5] R. Setiono and E. H. Liu. Symbolic representation of neural networks. *Computer*, 29(3):71-77, 1996.

[6] I. Taha and J. Ghosh. Evaluation and ordering of rules extracted from feedforward networks. *Proceedings of the IEEE International Conference on Neural Networks*, pages 221-226, 1997.

[7] C. A. Penã-Reyes and M. Sipper. Evolving *fuzzy* rules for breast cancer diagnosis. *Proceedings of 1998 International Symposium on Nonlinear Theory and Applications (NO*LTA'98), v. 2, pages. 369-372, Lausanne, 1998.

[8] R . S. Forsyth, D. D. Clarke and R. L. Wright. Overfitting Revisited: An InformationTheoretic Approach to Simplifying Discrimination Trees. *JETAI Journal of Experimental and Theoretical Artificial Intelligence*, 6,3, 1994.

[9] C. A. Penã-Reyes and M. Sipper. A fuzzy-genetic approach to breast cancer diagnosis. *Artificial intelligence in Medicine*, v. 17, v. 2, 1999

[10] J. Lukasiewicz. On 3-Valued Logic, *Ruch Filozoficzny*, 5, pages 169-171 (in Polish).

[11] H. S. Lopes, M. S. Coutinho and W. C. Lima. DoToR – um sistema especialista para apoio ao diagnóstico diferencial de dor toráxica. *Anais do III Fórum Nacional de Ciência e Tecnologia em Saúde / Congresso Brasileiro de Engenharia Biomédica*, Campos do Jordão, SP, pages 729-730, 1996.

[12] S. M. Weiss and C. A. Kulikowski. Computer systems that Learn - Classification and Prediction Methods from Statistics, Neural Nets, *Machine Learning, and Expert Systems*, Morgan Kaufmann, San Mateo, California, 1991.