

CLASSIFICAÇÃO DE DADOS RELATIVOS À CIRURGIA DE CÂNCER DE MAMA, UM COMPARATIVO ENTRE SOLUÇÃO POR REDES NEURAIIS E FUZZY.

Ademir C. Gabardo¹, Martín Pérez²

^{1,2}Universidade Tecnológica Federal do Paraná (UTFPR)
ademirgabardo@gmail.com, martin_perez@outlook.com

Resumo

Este artigo apresenta um comparativo entre duas técnicas para aprendizado de máquina baseadas em conceitos de inteligência artificial, utilizando Redes Neurais e controladores *Fuzzy* para classificar dados provenientes de pacientes submetidas à cirurgia de remoção de nódulos cancerígenas da(s) mama(s). O processo de classificação de dados pode ser simples ou complexo dependendo tanto do domínio do problema bem como da qualidade dos dados disponíveis. A acuidade dos resultados também pode variar de acordo com o domínio do problema, sendo que determinados casos podem admitir uma margem de erro maior ou menor. O objetivo deste trabalho é tanto demonstrar o processo de classificação de dados por meio de uma abordagem *Fuzzy* e se usando redes neurais, quanto demonstrar qual das técnicas obteve melhores resultados e com quais configurações.

Palavras-chave: Redes Neurais. *Fuzzy*. Câncer de mama. Mineração de dados.

Abstract

This paper presents a comparative study of two techniques for machine-learning concepts with artificial intelligence, using Neural Networks and Fuzzy controllers to classify data from patients undergoing surgery for removal of cancerous breast nodules. Process of data classification can be simple or complex depending on both the problem domain as well as the quality of available data. The accuracy of the results may also vary according to the problem domain, and some cases may permit a greater or lesser margin of error. The objective of this work is to demonstrate both the processes of data classification by a fuzzy approach and by using Neural networks, and also demonstrate how the techniques achieved the best results and with what settings.

Keywords: Neural Network. *Fuzzy*. Breast Cancer. Data-mining.

1. Introdução

Problemas de classificação estão presentes em diversos campos, desde a necessidade de separação de materiais até a identificação de padrões, reconhecimento de imagens e outros problemas de visão computacional.

Problemas de classificação de dados podem ser triviais quando os conjuntos são claramente separados e a regra de classificação é simples o suficiente para que não exista ambiguidade, no entanto isso nem sempre é o cenário que corresponde a problemas do mundo real.

Neste artigo propomos duas metodologias distintas para a classificação de dados provenientes de estudos realizados com pacientes que foram submetidas à cirurgia para a remoção de tumores causados pelo câncer de mama.

A compreensão de quais são os fatores importantes e quais atributos estão interligados a uma taxa de sobrevida ampla ou ao tempo de sobrevida dos pacientes é altamente relevante do ponto de vista clínico e consiste um problema computacional de difícil solução.

Objetivando identificar quais técnicas relacionadas à inteligência artificial e ao aprendizado de máquina obtém resultados mais apurados demandando de menores recursos

computacionais foi proposta uma abordagem utilizando Redes neurais artificiais em comparação a um classificador construído utilizando regras de classificação *Fuzzy*.

Os dados analisados são dados reais de pacientes submetidas à cirurgia e disponibilizados por meio da biblioteca *UCI Machine Learning*. São dados coletados pela equipe clínica do Dr. William H. Wolberg. Estes dados foram inseridos na biblioteca e publicados em 1992. As amostras são coletadas e analisadas periodicamente de acordo com os casos clínicos relatados. O banco de dados, portanto, reflete a ordem cronológica dos dados e recebe atualizações periódicas.

Este artigo está organizado da seguinte forma: As seções de um a seis fornecem uma breve fundamentação teórica das técnicas utilizadas. A seção sete detalha a metodologia utilizada. A oitava seção apresenta os resultados obtidos por ambos os métodos. E por fim, é feito um comparativo entre os resultados obtidos por cada um dos métodos.

2. Redes Neurais

Redes neurais artificiais são poderosos recursos computacionais (Yao, 1999) para o aprendizado de máquina, são sistemas computacionais inspirados no funcionamento do cérebro humano que emula a capacidade de aprendizado por meio da análise de dados de treinamento com o objetivo de abstrair uma porção de conhecimento (Krogh & Vedelsby, 1995). Nesta representação os neurônios são interligados por meio de nós. Para este trabalho é proposto uma metodologia capaz de classificar os dados de uma base conhecida a respeito da sobrevivência à cirurgia para remoção de tumores provocados pelo câncer de mama.

Por meio de Redes Neurais podemos estabelecer modelos baseados em métodos matemáticos capazes de representar complexas regiões de decisão (Mehrotra & Ranka, 1997). Tipicamente estas soluções são utilizadas quando existem muitos dados disponíveis, no entanto o problema não pode ser resolvido por meio de uma de uma regra simples ou de um algoritmo em tempo polinomial.

Redes Neurais emulam a capacidade de “aprender”, generalizar ou extrair regras automaticamente de conjuntos de dados complexos.

Redes Neurais podem ser aplicadas em diferentes tipos de tratamento de dados, e têm sido utilizadas em aplicações tais como: reconhecimento de voz, robótica, sistemas de transporte e roteamento, telecomunicações, eletrônica dentre muitas outras. Devido a sua vasta aplicabilidade as redes neurais tem ganhado espaço na comunidade científica para a resolução de problemas complexos onde o ser humano possui boa capacidade de resolução, como classificação de imagens, padrões, etc.

3. Redes MLP

Existem diversos tipos de redes neurais, dentre elas existem as redes MLP (*Multi Layer Perceptron*), cuja esquemática são representados pela figura 1 e figura 2, são redes neurais muito utilizadas por utilizarem poucos recursos computacionais e possibilitarem fácil implementação, ainda que não garantam um ponto de mínima global e possam demorar a convergir em certos casos, por exemplo, o aprendizado não linear (Arruda, 2012).

Sua arquitetura pode variar de acordo com o problema, tornando variável a necessidade de uso maior ou menor de neurônios ou de camadas.

As redes neurais recebem valores de forma análoga à como o cérebro recebe estímulos, estas funções são chamadas funções de ativação. Para isso, o software MATLAB fornece uma série de funções que são usualmente utilizadas com esse propósito. Entre as funções possíveis estão a tangente hiperbólica, a função logística e a função limiar.

Neste trabalho serão testadas diferentes arquiteturas aplicadas ao problema de classificação dos dados relativos à "Sobrevivência à cirurgia de câncer de mama", comparando o efeito da adição de neurônios e camadas na arquitetura para fins de aprendizado. Neste trabalho não se foi realizado nenhum tipo tratamento dos dados, o que pode ter influenciado negativamente o treinamento das redes.

As redes estudadas e escolhidas para a implementação deste trabalho foram redes MLP *Feed-foward backpropagation* com a função de aprendizado *Levenberg-Marquardt backpropagation*.

Todas as redes implementadas neste trabalho foram implementadas através da ferramenta MatLab 7.12.

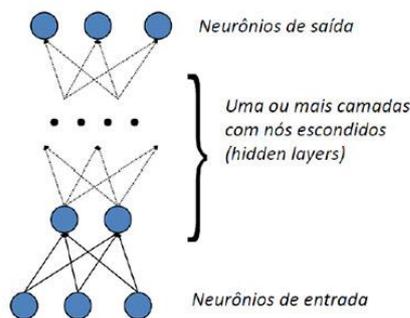


Figura 1 - Ilustração da arquitetura MLP.

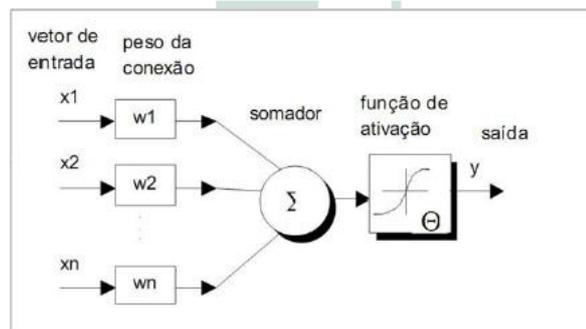


Figura 2 - Ilustração de um neurônio perceptron.

4. Lógica Fuzzy

A lógica Fuzzy ou lógica difusa é uma maneira de abstrair conceitos do mundo real que podem estar sujeitos à subjetividade transformando-as em um conjunto de regras matemáticas que seja passível de alimentar um algoritmo e inferir resultados com base nestes conjuntos de regras.

A lógica *Fuzzy* é uma extensão da lógica booleana que admite valores lógicos intermediários entre o FALSO (0) e o VERDADEIRO (1); por exemplo, o valor médio 'TALVEZ' (0,5). Isto significa que um valor lógico difuso é um valor qualquer no intervalo de valores entre 0 e 1. Este tipo de lógica engloba de certa forma conceitos estatísticos principalmente na área de inferência.

Deste modo abstraindo as informações para um dialeto compreensível e difuso que atenda as variações, incertezas e ambiguidades existentes nos problemas do mundo real. Um exemplo seria quantizar dimensões de esferas em conceitos como PEQUENO, MÉDIO ou GRANDE.

Existem diversas formas de criar os conjuntos de regras que irão servir de base para o algoritmo. Uma delas é contar com o auxílio de um especialista, uma ou mais pessoas

conhecedoras do universo do problema capazes de estabelecer quais métricas e valores são coerentes e pertinentes ao universo do problema a ser resolvido.

Outra maneira é utilizar dados previamente coletados sobre o problema a ser resolvido, obtendo como resultando em um conjunto de regras crível e coerentes a cerca do universo do problema. Isto é o que foi proposto no método de (Wang & Mendel, 1992).

5. Método de Wang e Mendel

Wang e Mendel propuseram em seu método que, um modo para obter os conjuntos de regras é a utilização de conjuntos de dados já conhecidos e obtidos por experimentação prévia para que então, a partir destes dados já conhecidos se obtenha os conjuntos de regras *Fuzzy* aplicáveis ao problema a resolver.

Esta metodologia é à base do algoritmo proposto por (Wang & Mendel, 1992) o qual, a partir de dados de treinamento gera um conjunto de regras *Fuzzy* probabilísticas, não descartando a informação inerente a dados redundantes ou conflitantes. Então o conhecimento de um especialista poderá neste caso ser útil para remover as ambiguidades e estabelecer critérios que por ventura não estejam presentes nos dados de treinamento.

6. Modelo de um controlador Fuzzy

O projeto de um controlador *Fuzzy* consiste em:

- Definir os universos de discurso das variáveis linguísticas (Zadeh, 1975) do sistema, erro, variação do erro e variação da saída do controlador (discretização ou níveis de quantização);
- Definição do número de termos primários e graus de pertinência dos conjuntos difusos que representam cada termo;
- Determinação das regras que formam o algoritmo de controle;

A figura 3 apresenta o diagrama de blocos para o funcionamento do mecanismo *Fuzzy*.

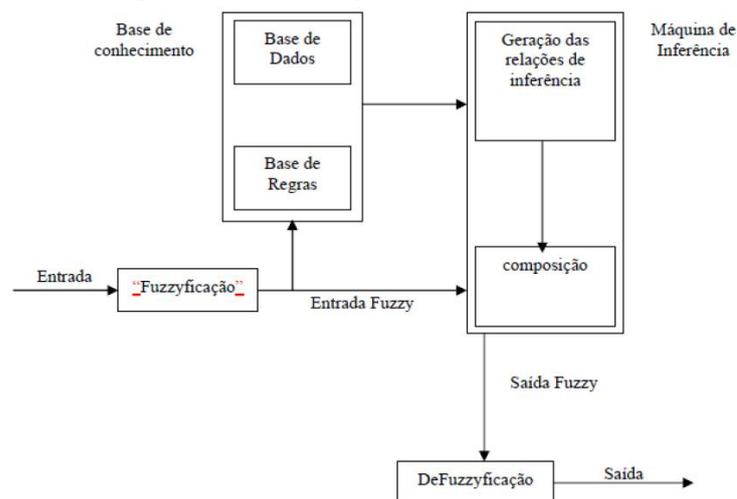


Figura 3 - Ilustração de um modelo de mecanismo *Fuzzy*.

7. Metodologia

O problema a ser resolvido proposto por este artigo é a classificação de uma base de dados composta de vários casos onde os seguintes atributos fazem parte do escopo do problema.

- Idade
- Ano
- Nódulos
- Classe

Estes dados representam a taxa de sobrevivida relacionada à cirurgia de câncer de mama em um determinado intervalo de tempo, sendo que as variáveis representam as seguintes informações: *idade* representa a idade em que a paciente realizou a cirurgia, *ano* representa o ano em que foi realizada a cirurgia, *nódulos* é a quantidade de nódulos localizados na região axilar e por fim, **sobrevida** que é o resultado obtido, sendo que, quando este valor é igual a 1 a probabilidade é de que a paciente tenha uma sobrevivida de 5 anos ou mais. Se o valor para sobrevivida for igual a 2 a probabilidade é de que a paciente tenha uma sobrevivida de até cinco anos.

7.1 Fuzzy

Para gerar as regras e também para efetuar os testes os seguintes parâmetros foram adotados para as variáveis:

A variável linguística de entrada **idade** foi dividida em três segmentos trapezoidais com as faixas de valor indicadas na tabela 1.

| <i>Faixa etária</i> | <i>De</i> | <i>Até</i> |
|---------------------|-----------|------------|
| Jovem | 10 | 40 |
| Adulto | 30 | 60 |
| Sênior | 50 | 80 |

Tabela 1 – Tabela de valores para a variável idade.

Estes valores resultam nos trapézios ilustrados na figura 4 que representam a entrada idade.



Figura 4 - Trapézios resultantes da variável idade.

A variável linguística de entrada **ano** que representa o ano em que foi realizada a cirurgia foi dividida em três segmentos trapezoidais com as seguintes faixas de valor indicadas na tabela 2.

| <i>Década</i> | <i>De</i> | <i>Até</i> |
|---------------|-----------|------------|
| Início | 1955 | 1965 |
| Meio | 1960 | 1970 |
| Final | 1965 | 1975 |

Tabela 2 – Tabela de valores para variável ano.

Estes valores resultam nos trapézios ilustrados na figura 5 que representam a entrada ano.

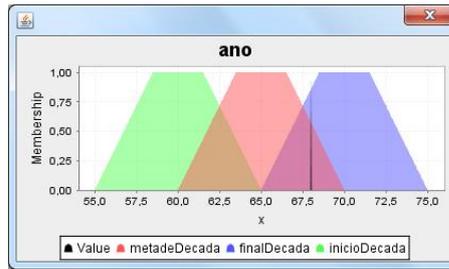


Figura 5 - Trapézios resultantes da variável ano.

A variável linguística **nódulos** foi dividida em três segmentos triangulares com as faixas de valores indicadas na tabela 3.

| <i>Número de Nódulos</i> | <i>De</i> | <i>Até</i> |
|--------------------------|-----------|------------|
| Poucos | 0 | 10 |
| Moderados | 10 | 20 |
| Muitos | 10 | 45 |

Tabela 3 – Tabela de valores para variável nódulos.

A figura 6 ilustra os triângulos que representam a variável de entrada nódulos.

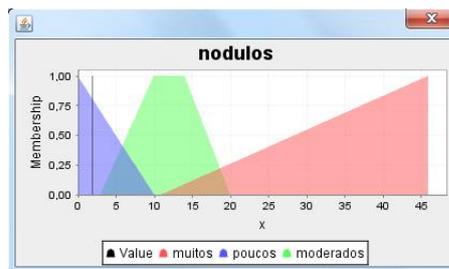


Figura 6 - Trapézios resultantes da variável nódulos.

E por fim, temos a variável de saída **sobrevida** que indica a qual classe resultara uma determinada configuração de variáveis de entrada. Esta variável trapezoidal foi definida com as seguintes faixas de valor conforme indicado na tabela 4.

| <i>Sobrevida</i> | <i>De</i> | <i>Até</i> |
|------------------|-----------|------------|
| Longa | 0 | 1 |
| Curta | 1 | 2 |

Tabela 4 – Variável de saída sobrevida.

A figura 7 ilustra os trapézios que representam a variável de saída sobrevida.



Figura 7 - Trapézios resultantes da variável de saída sobrevida.

Para a execução do trabalho foram fornecidos dois conjuntos de dados, uma base para treinamento de onde foram extraídos os conjuntos de regras. Na tabela de treinamento estão listados 244 casos que resultarão em 244 regras geradas por meio do algoritmo proposto por

(Wang & Mendel, 1992). Essas regras representam o conhecimento pré-existente a cerca do problema a ser resolvido.

Contudo, antes de aplicar as regras obtidas por meio do algoritmo é necessário proceder às desambiguações e remover as regras duplicadas para evitar a inconsistência dos resultados gerados. Um exemplo de ambiguidade pode ser resultado de dados de treinamento conflitantes, poderá então ser utilizado o apoio de um especialista para proceder a desambiguação das regras. Ou poderá ser escolhido um critério de pesos para as regras, sendo que, aquelas que ocorrerem um maior número de vezes terão prioridade. Depois de proceder as desambiguações e remover as regras duplicadas restaram 27 regras. Estas foram utilizadas para obtenção dos resultados junto aos dados de testes.

7.1 Redes Neurais MLP

Para criação das redes neurais foi se utilizada uma abordagem incremental para o número de neurônios e camadas, criando diversas redes MLP *Feed-foward backpropagation*, de forma de que redes com diferentes arquiteturas foram criadas, se aumentado o número de neurônios nas camadas ocultas e o número de camadas ocultas durante os testes.

Para os testes com redes de duas camadas (uma camada oculta) se foram testadas redes com 10, 20, 50, 100, 500 e 1000 neurônios na camada oculta.

Para os testes com redes três camadas (duas camadas ocultas) se foram testadas redes com 20, 50, 100 e 100 neurônios na primeira camada oculta e 20, 50, 50 e 100 neurônios na segunda camada oculta.

Para os testes com redes de quatro camadas (três camadas ocultas) se foram testadas redes com 20, 50, e 100 neurônios na primeira camada oculta e 20, 50, e 50 neurônios na segunda camada oculta, e 20, 50, e 50 neurônios na terceira camada oculta.

Todas as redes implementadas tiveram apenas um neurônio na camada de saída (classificador). Nenhum algoritmo de poda foi utilizado para se determinar o conjunto mínimo de neurônios necessários para a implementação de cada arquitetura.

Para cada arquitetura foram executados vinte treinamentos distintos. Foram avaliados os resultados de cada arquitetura para diferentes conjuntos de dados (teste e treinamento). Os valores de Idade, Ano e Nódulos foram os parâmetros de entrada dos treinamentos, sendo associados a uma classe resposta no treinamento. Os parâmetros avaliados foram o erro quadrático médio e percentual de acerto para cada conjunto.

8. Resultados obtidos

8.1 Fuzzy

Após a obtenção das regras de classificação pelo método Wang Mendel foi utilizada a biblioteca *JFuzzy* para implementação do código Java que recebe este conjunto de regras resultado das desambiguações e as mesmas configurações de variáveis de entrada e saída.

Com base nestas regras são efetivamente classificados os dados. O método utilizado para a saída é o COG (*Center of Gravity*). Com estas configurações de entrada e as regras utilizadas o índice de acerto em relação aos 61 casos estudados a partir dos dados de teste foi de 86,89%, sendo 53 acertos e 8 erros.

8.2 Redes Neurais

Após avaliar diversas arquiteturas possíveis com diferentes números de neurônios e camadas ocultas, os resultados obtidos nas diversas etapas das experiências foram submetidos aos seguintes critérios de análise.

- **MSE:** Erro quadrático médio da função para o conjunto total de dados de teste e de treinamento. Esta métrica considera o resultado contínuo da saída da rede neural sem nenhum tratamento.
- **Percentual de Acerto:** Divisão do número de acertos da saída da rede neural para os conjuntos de teste e de treinamento. Neste caso a saída da rede neural foi arredondada para corresponder a uma das duas classes pré-estabelecidas no problema.

As tabelas de 5 a 10 mostram os resultados encontrados para as diferentes arquiteturas implementadas neste trabalho.

| <i>MSE\Neurônios</i> | 10 | 20 | 50 | 100 | 500 | 1000 |
|----------------------|--------|--------|--------|--------|--------|--------|
| Todos os dados | 0,1539 | 0,1644 | 0,1501 | 0,2310 | 0,1572 | 0,2081 |
| Teste | 0,1647 | 0,1998 | 0,2151 | 0,2493 | 0,1950 | 0,1992 |
| Treino | 0,1512 | 0,1555 | 0,1339 | 0,2264 | 0,1477 | 0,2103 |

Tabela 5 – MSE para MLP de 2 camadas

| <i>Acerto\Neurônios</i> | 10 | 20 | 50 | 100 | 500 | 1000 |
|-------------------------|--------|--------|--------|--------|--------|--------|
| Teste | 0,8032 | 0,7377 | 0,7049 | 0,7540 | 0,7868 | 0,7868 |
| Treino | 0,7918 | 0,8000 | 0,8408 | 0,7632 | 0,8367 | 0,7836 |

Tabela 6 – Proporção de Acertos para MLP de 2 camadas

| <i>MSE\Neurônios</i> | 10 | 20 | 50 | 100 | 500 | 1000 |
|----------------------|--------|--------|--------|--------|--------|--------|
| Todos os dados | 0,1539 | 0,1644 | 0,1501 | 0,2310 | 0,1572 | 0,2081 |
| Teste | 0,1647 | 0,1998 | 0,2151 | 0,2493 | 0,1950 | 0,1992 |
| Treino | 0,1512 | 0,1555 | 0,1339 | 0,2264 | 0,1477 | 0,2103 |

Tabela 7 – MSE para MLP de 3 camadas

| <i>Acerto\Neurônios</i> | 20-20-1 | 50-50-1 | 100-50-1 | 100-100-1 |
|-------------------------|---------|---------|----------|-----------|
| Teste | 0,7540 | 0,7868 | 0,7377 | 0,8032 |
| Treino | 0,7591 | 0,8285 | 0,8285 | 0,8326 |

Tabela 8 – Proporção de Acertos para MLP de 3 camadas

| <i>MSE\Neurônios</i> | 20-20-1 | 50-50-1 | 100-100-1 |
|----------------------|---------|---------|-----------|
| Todos os dados | 0,1522 | 0,1658 | 0,1766 |
| Teste | 0,1922 | 0,1527 | 0,1998 |
| Treino | 0,1422 | 0,1690 | 0,1998 |

Tabela 9 – MSE para MLP de 4 camadas

| <i>Acerto\Neurônios</i> | 20-20-20-1 | 50-50-50-1 | 100-50-50-1 |
|-------------------------|------------|------------|-------------|
| Teste | 0,7377 | 0,7704 | 0,7377 |
| Treino | 0,8040 | 0,7755 | 0,7673 |

Tabela 10 – Proporção de Acertos para MLP de 4 camadas

Os resultados obtidos por meio do uso de redes neurais mostram que maior número de neurônios não implica necessariamente em melhor resultado (algoritmos de poda podem ser aplicados pra determinar o número ótimo de neurônios). Além disso, foi possível constatar que maior número de camadas não implica necessariamente em melhores resultados.

Outro fator a ser levado em consideração é que o sucesso do treinamento está relacionado à qualidade dos dados (sempre que possível os dados devem ser normalizados).

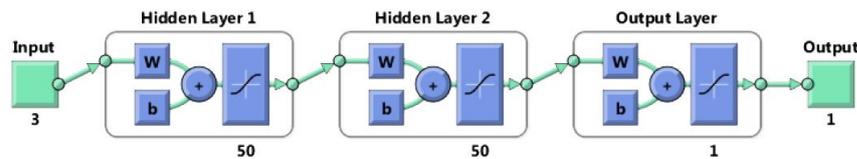


Figura 8 - Arquitetura da rede de maior sucesso entre as testadas.

9. Experimentos com RBF

Para fins didáticos estabeleceu-se uma comparação dos resultados obtidos com uma rede de base radial (RBF). Esta rede foi implementada através do uso de duas camadas, uma com função gaussiana e outra com função linear. Os resultados demonstraram-se muito superiores para os dados do conjunto de treinamento conforme pode ser visto na tabela 11, porém não foram eficientes para o conjunto de teste, possivelmente devido ao não tratamento do conjunto de dados.

| | | |
|-------------------------|--------|--------|
| <i>MSE\Neurônios</i> | 100 | 254 |
| Todos os dados | 0,0795 | 0,1025 |
| Teste | 0,3024 | 0,4732 |
| Treino | 0,0240 | 0,0102 |
| <i>Acerto\Neurônios</i> | 100 | 254 |
| Teste | 0,7213 | 0,7213 |
| Treino | 0,9180 | 0,9180 |

Tabela 11 – MSE e Acertos para rede neural RBF.

10. Conclusões.

Para as Redes Neurais observou-se que a qualidade dos dados, que são fatores tais como: ausência de ruídos, quantidade de dados para treinamento, complexidade dos dados que se deseja classificar tem profunda influência nos resultados, sendo que, a pura e simples configuração da rede com mais neurônios ou acréscimo de camadas não garantiu melhora significativa na precisão de classificação dos dados.

Para o método de classificação utilizando um controlador *Fuzzy* os ajustes feitos nas variáveis linguísticas foram os fatores com maior influência na geração das regras e, por conseguinte no resultado final da classificação.

Os melhores resultados foram obtidos utilizando-se a metodologia de classificação de dados por meio de um controlador *Fuzzy*, onde se obteve uma taxa de acertos maior que aquela observada nos resultados obtidos utilizando-se Redes Neurais.

Quanto ao consumo de recursos computacionais não se observaram grandes variações de tempo no processamento em ambas às técnicas aplicadas.

Como ponto forte das Redes Neurais pode-se destacar a possibilidade de ampliar o número de atributos sem a necessidade da troca de modelo, sendo que para o método que utiliza *Fuzzy* a vantagem foi a maior precisão dos resultados obtidos.

Como pontos fortes de ambos os métodos pode-se destacar a necessidade de ajustes de acordo com o conjunto de dados utilizados.

Utilizando ambas as técnicas foi possível obter resultados satisfatórios de classificação dos dados. Sendo que em ambos os métodos a porcentagem de acerto foi superior a setenta por cento o que caracteriza como bons métodos de classificação.

Referências

ARRUDA, V., (ed.). (2012). **RNA - notas de aulas.**

KROGH, A. & Vedelsby, J. (1995). **Neural Network Ensembles, Cross Validation, and Active Learning.** In Advances in Neural Information Processing Systems (pp. 231--238). MIT Press.

MEHROTRA, K., Mohan, C. K. & Ranka, S. (1997). **Elements of artificial neural networks.** the MIT Press.

MELO, L. G. d. (2011). **Sistemas Fuzzy Probabilísticos: Geração Automática De Regras E Defuzzificação Bayesiana.** Unpublished master's thesis, UTFPR - Unversidade Tecnologica Federal do Paraná.

WANG, L.-X. & Mendel, J. (1992). **Generating fuzzy rules by learning from examples.** IEEE Transactions on Systems, Man and Cybernetics 22, 1414 - 1427.

YAO, X. (1999). **Evolving Artificial Neural Networks.**

ZADEH, L. (1965). **Fuzzy Sets.** Information and Control.

ZADEH, L. A. (1975). **The Concept of a Linguistic Variable and its Application to Approximate Reasoning.** Journal of Information Science, 199.

