



Rede Neural Artificial aplicada ao prognóstico de câncer de próstata

Genilto Dallo¹ (IC-Voluntária/ UNICENTRO), Angelita Maria De Ré¹, Fábio
Hernandes¹, Sandro Rautenberg¹, Maria José de Paula Castanho²
(Orientadora), e-mail: geniltodallo@gmail.com

Universidade Estadual do Centro-Oeste/Departamento de Ciência da
Computação¹/Departamento de Matemática²/Guarapuava, PR.

Ciências Exatas e da Terra, Ciência da Computação

Palavras-chave: Câncer de Próstata, Prognóstico, Rede Neural Artificial
Perceptron Multicamadas.

Resumo:

O câncer de próstata é o sexto tipo de câncer mais comum no mundo, com cerca de 10% das ocorrências. Este trabalho investiga o desenvolvimento de uma Rede Neural Artificial *Perceptron* Multicamadas para auxiliar o médico na tomada de decisão quanto a realizar ou não a cirurgia, que deve ser feita apenas se o tumor estiver confinado na próstata. A Rede Neural Artificial é utilizada para classificar cada caso de câncer como confinado ou não-confinado a partir de dados de pacientes. Porém, o uso deste tipo de RNA para o prognóstico de câncer de próstata ainda não apresentou o comportamento esperado, motivando trabalhos futuros.

Introdução

O câncer de próstata é o sexto tipo de câncer mais comum no mundo e o mais prevalente em homens, representando cerca de 10% dos casos. No Brasil, no corrente ano, é estimado um risco de 54 casos novos a cada 100 mil homens [INCA, 2010]. Quando detectado precocemente a cura atinge 90% dos casos após a prostatectomia radical.

Neste contexto, uma importante decisão a ser tomada pelo médico é a realização ou não de cirurgia para retirada do tumor. Para tal decisão, é necessário detectar o estágio do câncer, pois se o tumor já atingiu estruturas adjacentes (câncer não-confinado), outras terapias são mais apropriadas. Os principais indicadores para determinar o estágio do câncer de próstata são: o nível de PSA (*Prostate Specific Antigen*) no sangue, o escore de Gleason da biópsia [CALVETE *et al*, 2003] e o exame clínico (exame digital retal e/ou ultrassom).

Como um mecanismo para auxiliar o médico em seu prognóstico, este trabalho investiga o desenvolvimento de uma Rede Neural Artificial *Perceptron* Multicamadas utilizando o algoritmo de aprendizado supervisionado *backpropagation* para classificar os pacientes com câncer confinado ou não-confinado na próstata. Considera-se como entradas as informações sobre os resultados dos exames PSA, Gleason e exame clínico, além da idade e cor do paciente.

Materiais e métodos

Segundo Freeman e Skapura, [1991], Redes Neurais Artificiais (RNA) que utilizam *backpropagation* são aplicadas como soluções apropriadas em problemas que necessitam abstrair generalizações de dados. Isso significa que, dado uma quantidade de dados de entrada, uma RNA tem a capacidade de agrupar dados similares em classes, propiciando sentido a respeito dos dados. Tal característica é explorada neste trabalho.

A RNA *Perceptron* Multicamadas (MLP), é uma rede constituída de uma ou mais camadas internas de neurônios. Os neurônios possuem laços de realimentação, ou seja, a saída de um neurônio pode ser entrada para outro de uma camada. Neste trabalho, dentre os vários algoritmos existentes para treinar as MLP, foi escolhido o algoritmo *backpropagation* [HAYKIN, 2001]. Este algoritmo de aprendizado supervisionado é baseado na correção de erros e tem por objetivo encontrar o erro mínimo global, maximizando o poder de generalização da rede. A função de ativação utilizada por este algoritmo de aprendizado foi a sigmoideal [RUSSEL e NORVIG, 1995].

Os parâmetros da MLP investigada no trabalho são: número de neurônios de entrada; número de neurônios na camada intermediária; número de neurônios na camada de saída; épocas de treinamento e taxa de aprendizado. Nas simulações, o conjunto de neurônios na camada de entrada era composto por 12 elementos de processamento e na camada de saída apenas um neurônio (Figura 1). Os parâmetros da rede testados foram: taxa de aprendizado {0,1; 0,2; 0,3; 0,4; 0,5; 0,7}; número de neurônios da camada intermediária {14, 15, 18, 20, 22, 24, 32}; e épocas de treinamento {100, 1000, 5000 e 10000}.

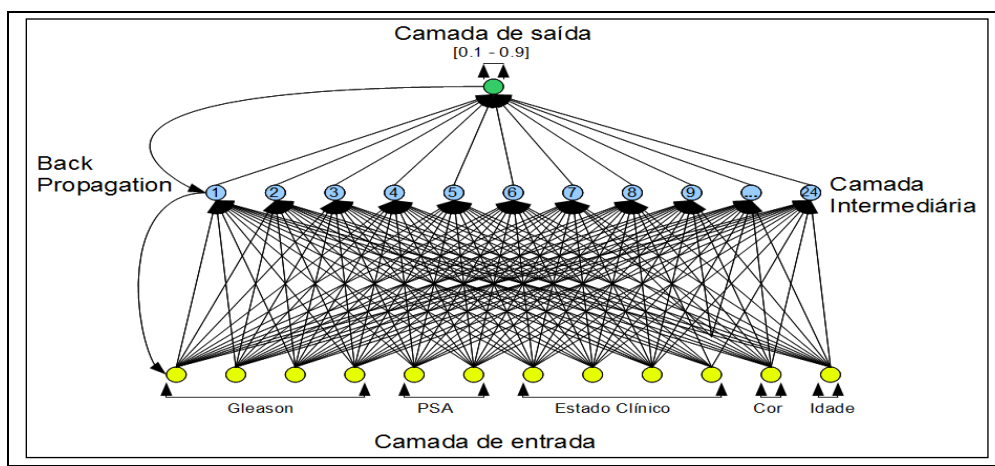


Figura 1. MLP desenvolvida

A base de dados¹ utilizada é composta por dados pré-cirúrgicos de 267 pacientes, além da classificação do câncer em confinado na próstata ou não. Do conjunto de dados disponíveis, 187 casos foram utilizados para o

¹ Hospital das Clínicas - UNICAMP

treinamento da rede (aproximadamente 70% da base de dados) e 80 casos para testes (30%), escolhidos aleatoriamente dentro de cada classe.

A camada de entrada da RNA implementada foi configurada para obter os seguintes dados da base: PSA com 2 neurônios, Gleason com 4 neurônios, Estado clínico com 4 neurônios, Cor e Idade com um neurônio cada. A Tabela 1 exemplifica a base configurada para a entrada da rede.

Tabela1. Exemplo de dados configurados para entrada da rede.

| Gleason | | | | PSA | | Estado | | | | Diagnóstico | | Cor | Idade |
|---------|---|---|---|--------|--------|--------|---|---|---|-------------|---|-----|-------|
| 1 | 0 | 0 | 0 | 0 | 0.5985 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 61 |
| 0 | 0 | 0 | 1 | 0.9989 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 46 |
| 0 | 1 | 0 | 0 | 0 | 0.4359 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 64 |
| 1 | 0 | 0 | 0 | 0 | 0.2854 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 55 |

Para o treinamento, os dados são inseridos na camada de entrada e são propagados até a camada de saída. Os valores de saída são comparados com o resultado desejado (0 ou 1). Ou seja, se é confinado (1) ou não confinado (0), e um sinal de erro é computado e transmitido da camada de saída para cada neurônio da camada intermediária que contribuem diretamente com aquela saída, ajustando assim os valores (pesos das sinapses) de cada neurônio.

A classificação ou teste é feito inserindo os dados na camada de entrada e comparando o diagnóstico da base com a saída da rede. Com base nos valores de saída, um ponto médio é escolhido como ponto de corte para a classificação.

Um protótipo de sistema foi desenvolvido na linguagem Java utilizando a IDE NetBeans para execução dos treinamentos e testes, possibilitando a parametrização de diversas topologias de RNA.

Resultados e Discussão

Como mencionado, foram realizadas simulações de topologias de RNA, com diferentes valores para taxa de aprendizado, número de neurônios na camada interna, e épocas de treinamento. Cada topologia simulada foi executada 20 vezes, com o objetivo de transpor a desvantagem da suscetibilidade dos pesos iniciais e investigar a estabilidade da taxa de acerto da RNA implementada.

Durante as simulações, com os parâmetros configurados em: taxa de aprendizado de 0,4; 24 neurônios na camada intermediária; e 1000 épocas, encontrou-se uma topologia de RNA com a taxa de acerto entre 61% e 96%, (intervalo que comporta a melhor RNA encontrada).

Porém, para a configuração da RNA supracitada, observando a totalidade dos testes realizados, é possível perceber que resultados significativos não ocorreram com a frequência desejada. Na Figura 2, a área sob a curva ROC [ZWEIG e CAMPBELL,1993] referente a RNA com taxa média de acerto, na configuração mencionada, é 0,67 (IC: 0,56 – 0,77) Na

literatura, outras abordagens apresentam a área em torno de 0,78 [AUGUSTIN *et al*, 2004].

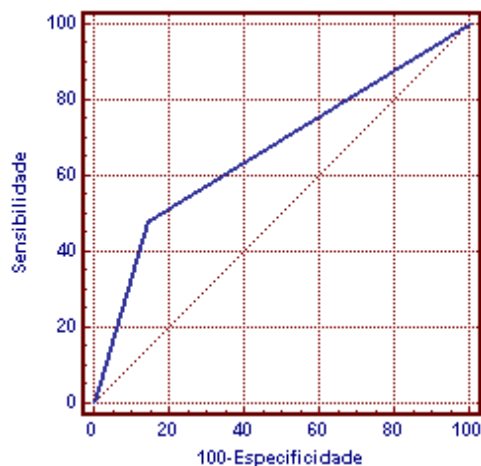


Figura 2. Curva ROC

Conclusões

Por este estudo, pode-se concluir que o uso deste tipo de RNA para o prognóstico de câncer de próstata não apresentou o comportamento esperado. Entretanto, o fato de ter ocorrido uma configuração com 96% de acerto motiva a continuidade da pesquisa. Como trabalhos futuros, pretende-se avaliar a eficiência da RNA, perante testes mais minuciosos, considerando outras partições da base de dados e, também, propor uma modificação na regra delta generalizada, incluindo o *momentum*, para identificar o efeito das mudanças na alteração dos pesos.

Referências

- Augustin, H. Eggert, T. Wenske, S. Karakiewicz, P.I. Palisaar, J. Daghofer, F. Huland, H. Graefen, M. Comparison of Accuracy between the Partin Tables of 1997 and 2001 to predict final pathological stage in clinically localized prostate cancer, *J. Urol.*, 171 (2004), 177-81.
- Calvete A. C. et al. Avaliação da Extensão da Neoplasia em Câncer da Próstata: Valor do PSA, da Percentagem de Fragmentos Positivos e da Escala de Gleason, *Rev. Assoc. Med. Bras.* 2003, 49, 250-254.
- Freeman, J A. e Skapura, D M., *Neural networks : algorithms, applications, and programming techniques*, Addison Wesley, 1991.
- Haykin, S. S. *Redes Neurais Princípios e Prática*. Bookman Companhia, 2001
- INCA – Instituto Nacional do Câncer, Ministério da Saúde, disponível em <http://www.inca.gov.br/estimativa/2010>, acesso em 26.04.10.
- Russell, S J e Norvig, P. *Artificial Intelligence: a modern approach*. New Jersey: Prentice-Hall Inc., 1995.



Zweig, M. H.; Campbell, G. “Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine,” *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1993.