

APLICAÇÃO DE UMA METODOLOGIA COMPUTACIONAL INTELIGENTE NO DIAGNÓSTICO DE LESÕES CANCERÍGENAS

Juliana Oliveira Ferreira

Mestrando em Ciência da Computação pela Universidade Federal de Minas Gerais na Área de Banco de Dados/Mineração de Dados. Bacharel em Ciência da Computação pela Universidade Federal de Alfenas. juolivfer@gmail.com.

Humberto César Brandão de Oliveira

Doutorando em Ciência da Computação pela Universidade Federal de Minas Gerais, na área de Otimização Combinatória. Mestre em Ciência da Computação pela Universidade Federal de Pernambuco, na área de Inteligência Artificial. Bacharel em Ciência da Computação pela Universidade Federal de Lavras. Professor Adjunto da Universidade Federal de Alfenas. humberto@bcc.unifal-mg.edu.br.

Maria Regina Martinez

Doutora e Mestre em Ciências pela Universidade de São Paulo. Graduação em Enfermagem pela Universidade de São Paulo. Atualmente é Professora Adjunta da Universidade Federal de Alfenas atuando em docência, pesquisa e extensão na área de Administração aplicada à Enfermagem. martinez@unifal-mg.edu.br.

RESUMO

Este trabalho apresenta um Sistema Inteligente utilizado para indicar, através de técnicas de Inteligência Artificial, se células da mama apresentam tumores malignos.

A técnica apresentada corresponde a uma adaptação da Programação Genética e consiste em localizar expressões matemáticas que, ao serem processadas com características numéricas da lesão a ser analisada, possibilitam a classificação.

Testes foram realizados em um banco de dados de células reais e através dos resultados podemos concluir que o método proposto pode ser uma boa ferramenta para auxiliar os Patologistas no diagnóstico de lesões cancerígenas, principalmente quando o profissional ainda não adquiriu muita experiência. A técnica também foi capaz de preservar o conhecimento de profissionais experientes e selecionar as características importantes para realizar a classificação.

PALAVRAS-CHAVE: Sistemas Inteligentes, Neoplasias Mamárias, Câncer de Mama, Classificadores.

APPLICATION OF AN INTELLIGENT COMPUTATIONAL METHODS IN DIAGNOSIS OF CANCEROUS LESIONS

ABSTRACT

This paper presents an Intelligent System used to indicate, through Artificial Intelligence techniques, cells were present in breast malignancies.

This technique represents an adaptation of Genetic Programming that find mathematical expressions. When this one is processed with numerical characteristics of the lesion, the cell can be classified.

Tests were performed on a database of real cells and through the results we can conclude that the proposed method can be a good tool to assist pathologists in the diagnosis of

cancerous lesions, especially when the professional has no good experience. The technique was also able to preserve the knowledge of experienced professionals and select the important features to perform classification.

KEYWORDS: Intelligent Systems, Mammary Neoplasm, Breast Cancer, Classification.

APLICAÇÃO DE UMA METODOLOGIA COMPUTACIONAL INTELIGENTE NO DIAGNÓSTICO DE LESÕES CANCERÍGENAS

INTRODUÇÃO

Com o envelhecimento e maior expectativa de vida da população há um aumento consistente da prevalência de doenças crônicas, destacando-se o câncer. Principalmente em países de médio e baixo desenvolvimento, onde se enquadra o Brasil, a prevalência de câncer na população torna-se um problema de saúde pública que exige vigilância e condutas de prevenção e tratamento efetivas. A estimativa para 2010 da ocorrência de câncer no Brasil é maior de 480 mil casos novos¹.

O correto diagnóstico e estadiamento clínico dos tumores cancerígenos fornecem apoio decisório nas condutas terapêuticas e aconselhamento do paciente, sendo utilizados diferentes sistemas de classificação^{2,3}.

Os sistemas de classificação de tumores são detalhados e complexos e, normalmente, de difícil utilização objetiva pelos patologistas, oferecendo limitações no seu uso^{3,4}.

Sistemas computacionais inteligentes têm sido desenvolvidos como ferramenta de apoio decisório em diferentes áreas da saúde⁵⁻⁶, e podem contribuir com o patologista para uma classificação mais objetiva e uniforme dos tumores analisados, minimizando as limitações de uso impostas pelos sistemas de classificação, acelerando o trabalho desse profissional.

O principal objetivo deste trabalho foi desenvolver um sistema computacional inteligente capaz de realizar a classificação das células tumorais, informando a sua malignidade.

O sistema computacional se baseou no desenvolvimento e adaptação da técnica denominada Programação Genética⁷. Com o treinamento do sistema foram localizadas expressões matemáticas que, ao serem processadas com características numéricas da lesão tumoral a ser analisada, possibilitaram a classificação da malignidade da lesão.

MÉTODOS

Para realização dos experimentos utilizou-se de uma base de dados reais e bem explorada pela literatura conhecida como Proben1, publicada no Relatório Técnico de Prechelt¹² e encontrada no seguinte endereço eletrônico: <ftp://ftp.ira.uka.de/pub/neuron/proben1.tar.gz>. Foram obtidos dados relativos a características de células tumorais de câncer de mama obtidas por imagens digitalizadas, bem como sua classificação de malignidade. Para cada uma das 699 amostras da base, são informadas 9 características das células, sendo estas: espessura da massa celular, uniformidade do tamanho da célula, uniformidade do formato da célula, adesão marginal, tamanho de uma célula epitelial, núcleo vazio, cromatina branda, nucléolo normal e mitose. Além de indicar se a mesma possui tumor maligno ou benigno. Das 699 células, 65% possuem tumor benigno. Este banco de dados é composto por 3 sub-bases distintas (cancer1, cancer2 e cancer3), cada uma com 699 informações de células.

A Técnica de Classificação proposta por este trabalho consistiu em uma adaptação da Programação Genética^{7,8}, onde o classificador busca expressões aritméticas capazes de realizar a categorização de lesões cancerígenas em diferentes classes.

As diferentes fórmulas matemáticas representáveis nesta metodologia foram definidas em uma linguagem específica livre de contexto^{9,10}, e a representação computacional foi feita por Árvores Sintáticas^{4,5} como o exemplo mostrado na Figura 1.

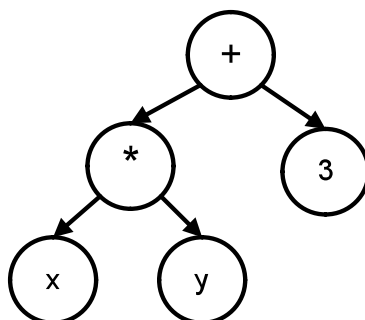


Figura 1 – Representação da função $g(x,y) = x*y+3$.

Cada árvore armazena uma expressão que contém funções matemáticas (soma, subtração e multiplicação) constantes e características do objeto a ser identificado (mapeadas em números reais). A busca por expressões bem adaptadas ao problema é feita através de pequenas alterações nas árvores já encontradas e cabe ao sistema determinar a melhor expressão baseando-se no histórico já conhecido na base de dados Proben1. Sempre na raiz da expressão do sistema existe uma função sigmoïdal, capaz de fornecer uma classificação binária (0 – tumor benigno, 1 – tumor maligno).

A métrica utilizada para avaliação das expressões foi quantidade de acertos e a Soma dos Erros Quadráticos (Sum Square Error-SSE), análise amplamente utilizada na arquitetura de diferentes Redes Neurais Artificiais¹¹. A métrica SSE é descrita na Equação 1.

equação (1)

O sistema foi programado para fazer três tipos diferentes de análise para selecionar a melhor expressão: (1) escolher a expressão com menor SSE (ST1), (2) escolher a expressão com maior taxa de acerto (ST2) ou (3) escolher a expressão com menor SSE e maior taxa de acerto (ST3).

Para treinamento do sistema e realização dos testes experimentais foram utilizados 699 amostras obtidas da base Proben1. Para cada uma das amostras foram informadas 9 características, além de se indicar a sua classificação como tumor maligno ou benigno. Uma característica deste trabalho é a possibilidade do sistema encontrar fórmulas que não utilizam todas as características disponibilizadas na base de dados do hospital Norte-Americano Madison da Universidade de Wisconsin.

O sistema inicia o processo com fórmulas aleatórias. A princípio, tais fórmulas possuem alta probabilidade de não conseguirem de fato classificar o tipo de tumor das células do banco de dados. Com o decorrer do processo de otimização, as fórmulas vão sofrendo

pequenas alterações, e através da métrica SSE é possível avaliar se elas são mais adaptadas a classificação correta das às células do banco de dados. Com menos de um minuto de simulação é possível encontrar células que possuem taxa de acerto acima de 95%. Todo o processo experimental pode ser encontrado no trabalho de Ferreira¹³.

RESULTADOS E DISCUSSÃO

O passo inicial do Sistema consistiu na entrada com dados previamente conhecidos, ou seja, fornecer dados de algumas lesões cancerígenas informando as suas respectivas classes (maligno ou benigno). O objetivo desta fase é o treinamento do Sistema Inteligente.

Logo após a localização das melhores expressões capazes de classificar o câncer através de características básicas da célula, o sistema estará pronto para realizar novas classificações (de células que ele desconhece). Neste ponto, basta informar os dados da lesão ao sistema e este retorna o diagnóstico.

Para realização dos experimentos foram efetuadas 500 execuções, sendo o resultado deste apresentado na Figura 2. No eixo das abscissas a Figura 2 apresenta o tipo de algoritmo aplicado (variando sua métrica de qualidade), e no eixo das ordenadas temos o percentual de acerto do algoritmo testado. O tipo de gráfico da Figura 2 é o boxplot.

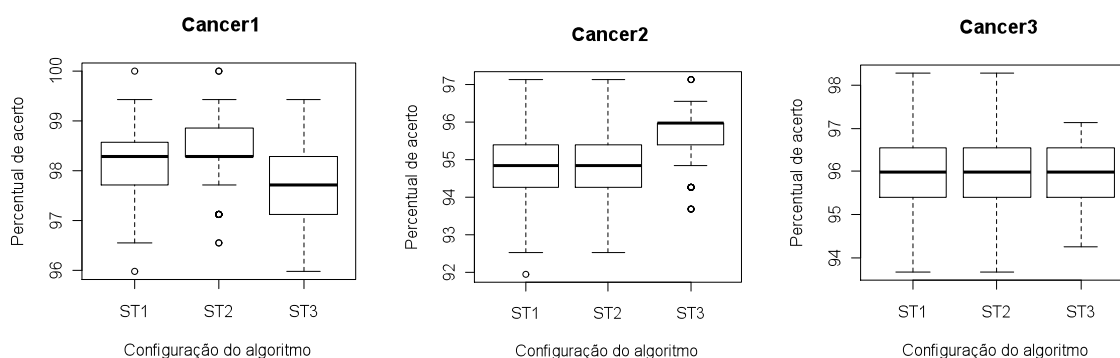


Figura 2 – Resultado da classificação para as três bases de dados.

A Tabela 1 apresenta a média (MD), o desvio padrão (DP), o melhor resultado alcançado (M), e a quantidade de componentes utilizadas (C) pela melhor fórmula encontrada pelo sistema para cada base de dados, de um total de 9 características da célula.

Tabela 1: Resultados obtidos pelo classificador para as três configurações.

Configuração	Cancer1				Cancer2				Cancer3			
	MD	DP	M	C	MD	DP	M	C	MD	DP	M	C
ST1	98,16	0,62	100	9	94,72	0,86	97,13	9	96,17	0,76	98,28	9
ST2	98,48	0,59	100	6	94,74	0,82	97,13	7	95,98	0,70	98,28	9
ST3	97,63	0,56	99,43	9	95,72	0,60	97,13	8	95,83	0,62	97,7	9

Nota-se que os resultados obtidos pela pesquisa são promissores, principalmente na base Câncer 1, onde o melhor resultado conseguiu acertar todos os diagnósticos feitos. Outro fato importante a ser observado é a capacidade de se selecionar apenas algumas características da célula para efetuar a classificação efetiva. Este fato é importante na redução de custos de futuras classificações.

CONCLUSÕES

O objetivo central desta pesquisa foi o desenvolvimento de um Sistema Inteligente capaz de auxiliar patologistas no diagnóstico rápido e assertivo de achados histopatológicos de lesões cancerígenas com o intuito de facilitar a escolha do tratamento mais adequado, aumentando assim as chances de sucesso no tratamento do paciente.

A técnica basicamente tenta encontrar expressões matemática que sejam capazes de classificar essas lesões em suas diferentes classes, e pode ser adaptada para outros tipos de classificações.

Os resultados foram satisfatórios e mostraram que o Sistema pode ajudar consideravelmente os especialistas a realizar este diagnóstico, podendo ser usado como uma forma de capturar e organizar o conhecimento e o talento de profissionais experientes, mantendo assim, este conhecimento na organização e até mesmo o disponibilizando para ajudar patologistas menos experientes a determinem o diagnóstico com mais confiabilidade. O tempo gasto pelo sistema é consideravelmente pequeno, sendo considerado apenas o tempo gasto na coleta dos dados. O Sistema conseguiu também selecionar automaticamente as características mais importantes para a classificação podendo-se, desta forma, diminuir o custo na coleta de informações para futuras classificações.

O sistema computacional mostrou-se adaptável para a realização de outros tipos de classificações, funcionando como uma forma de preservar, aproveitar e organizar o talento e a experiência de especialistas. Uma das vantagens do sistema foi a capacidade de armazenar o conhecimento de mais de um profissional, diminuindo o custo e o tempo gastos em casos que o patologista considere necessário a análise de mais de um especialista.

REFERÊNCIAS BIBLIOGRÁFICAS

1. INSTITUTO NACIONAL DE CÂNCER (Brasil). Estimativa 2010: incidência de câncer no Brasil. Rio de Janeiro: INCA; 2009.
2. PATEL, S.G. LYDIATT, W.M. Staging of head and neck cancers: is it time to change the balance between the ideal and the practical? *J Surg Oncol* 2008; 97:653-57.
3. LOURENÇO, S.Q.C. SCHUELER, A. CAMICASCA, D.R. LINDENBLATT, R.C. e BERNARDO, V.G. Classificações Histopatológicas para o Carcinoma de Células Escamosas da Cavidade Oral: Revisão de Sistemas Propostos. *Revista Brasileira de Cancerologia*. 2007; C43. 53.3-9 2007.
4. MANIKANTAN, K. SAYED, S.I. SYRIGOS, K.N. RHYS-EVANS, P. NUTTING, C.M. HARRINGTON, K.J. KAZI, R. Challenges for the future modifications of the TNM staging system for head and neck cancer. Case for a new computational model? *Cancer Treatment Reviews* 35 (2009) 639-44.
5. KROGH, A. What are artificial neural networks? *Nature Biotechnology* 2008; 26(2):195-7.

6. TOWNSEND, D. FRIZE, M. Complimentary artificial neural network approaches for prediction of events in the neonatal intensive care unit. IEEE Engineering in Medicine and Biology Society Conference Vancouver. IEEE Computer Society Press, 2008, pp. 20-24.
7. KOZA, J.R. Genetic Programming : On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA, 1992.
8. EIBEN, A.E. e SMITH, J.E. Introduction to Evolution to Evolutionary Computing. Natural Coputing Series, Mit Press. Springer. Berlin, 2003.
9. WHIGHAM, P.A. (1996), Grammatical bias for evolutionary learning. University New South Wales, Austrália.
10. WONG, M.L. e LEUNG, K.S. (2000), Data mining using grammar based genetic programming and applications. Kluwer Academic Publishers.
11. HAYKIN, S. Neural Networks: A comprehensive Foundation. 2nd Edition, Prentice Hall, 1998.
12. PRECHELT, L. Proben1: A set of Neural Network Benchmark Problems and Benchmarking Rules. Technical Report 21/94, Fakultät für Informatik, Universität Karlsruhe, 76128 Karlsruhe, Germany, September, 1994.
13. FERREIRA, J.O., OLIVEIRA, H.C.B., PAULA, M.M.V. Sigm-Tree: Otimização de Árvores Sintáticas Aplicadas à Classificação de Padrões. Simpósio Brasileiro de Pesquisa Operacional, 2009, Porto Seguro.