

ESTUDO COMPARATIVO DE ABORDAGENS INTELIGENTES PARA AUXÍLIO AO DIAGNÓSTICO DE MASSAS NODULARES EM IMAGENS MAMOGRÁFICAS

Antonio Gabriel Pereira de Andrade¹; Ana Lúcia Marreiros Maia²

1. Bolsista PROBIC/UEFS, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, email: gabriel.ecomp@gmail.com

2. Orientadora, Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, email: anamarreiros@gmail.com

PALAVRAS-CHAVE: classificador fuzzy, classificador bayesiano, câncer de mama.

Introdução

O câncer de mama é uma das doenças mais temidas entre as mulheres devido à sua frequência, alta taxa de mortalidade feminina em todo o mundo e, sobretudo, pelos seus efeitos psicológicos, que afetam a percepção da sexualidade e a própria imagem pessoal. Como suas causas ainda não são bem conhecidas, a prevenção é a única maneira de reduzir a mortalidade de mulheres com câncer de mama, pois a detecção precoce aumenta as chances de cura substancialmente [1, 2].

Por se tratar de um problema de extrema importância, muitas pesquisas vêm sendo desenvolvidas com o intuito de indicar uma abordagem computacional eficaz no auxílio da difícil tarefa de analisar estruturas relevantes em imagens mamográficas e diagnóstico de lesões, dando ao especialista um segundo parâmetro de análise para o diagnóstico final. O objetivo deste trabalho foi o desenvolvimento de um classificador fuzzy e um classificador bayesiano para operar eficientemente sobre uma base de dados contendo casos diagnosticados de câncer de mama.

Para a modelagem da base de dados e base de regras referentes ao classificador fuzzy, baseamo-nos em um trabalho desenvolvido por Pena-Reyes e Sipper [5], por se tratar de uma modelagem fuzzy de classificação aplicado sobre a mesma base de dados de Wisconsin[4] utilizada neste trabalho.

Além do sistema fuzzy, também foi desenvolvido um classificador bayesiano simples para operar sobre a base de Wisconsin. Esse classificador é baseado em modelos estatísticos de probabilidade e apresentou ótimos resultados quando comparado ao classificador fuzzy.

Material e Métodos

A metodologia utilizada no desenvolvimento da pesquisa seguiu as seguintes etapas:

- (1) Teórica, onde foram analisados os principais modelos de sistemas fuzzy e sua aplicação na classificação de padrões. Nesta etapa também foi feita uma revisão da literatura, a fim de compreender as metodologias empregadas em trabalhos relacionados.
- (2) Ferramental, onde foram modelados e implementados os classificadores.
- (3) Experimental, na qual foram realizados os experimentos necessários à verificação da aplicabilidade e eficácia dos classificadores fuzzy e bayesiano.
- (4) Testes e Análise dos Resultados, na qual os resultados obtidos da aplicação das abordagens propostas foram analisados e comparados com outras abordagens disponíveis na literatura;

A base de dados escolhida para o trabalho foi o Conjunto de Dados de Câncer de Mama da Universidade de Wisconsin (*Breast Cancer Wisconsin Diagnostic Data Set*) [3, 4]. A base de Wisconsin é uma base pública, e encontra-se disponível, para qualquer pesquisador interessado, no Repositório de Aprendizado de Máquina da Universidade da Califórnia (Irvine) [3]. Ela contém 699 amostras diagnosticadas, divididas em duas classes: 458 casos benignos (65,5%) e 241 casos malignos (34,5%). As amostras apresentam características computadas sobre imagens digitalizadas originárias de procedimentos de aspirações de agulha fina (fine needle aspirate - FNA), uma técnica menos invasiva que as biópsias tradicionais e com maior índice de acerto que imagens mamográficas. Cada amostra é constituída de 11 atributos numéricos, incluindo a classe ao qual ela pertence (maligno ou benigno) e valores representando as características do núcleo das células nas imagens.

Após a definição da base de dados de amostras, foi necessário definir os parâmetros do classificador fuzzy. Para isso, nos baseamos no trabalho desenvolvido por Pena-Reyes e Sipper [5] no qual eles propõem uma abordagem automatizada para gerar os melhores parâmetros para um classificador fuzzy atuando sobre a base de Wisconsin. Esses parâmetros consistem de:

- Variáveis fuzzy com dois rótulos: **Alto** e **Baixo**
- Funções de pertinência do tipo ortogonal.
- Operador fuzzy min-max.
- Defuzzificação por média ponderada.
- Saída binária: maligno (2) e benigno (4).

- Base de regras composta por apenas duas regras:
 - 1- **se** (*v1 é Baixo*) e (*v3 é Baixo*) e (*v7 é Baixo*) e (*v8 é Baixo*) **então** (*saída é Benigno*).
 - 2- **se** (*v1 é Baixo*) e (*v4 é Baixo*) e (*v6 é Baixo*) e (*v8 é Baixo*) e (*v9 é Baixo*) **então** (*saída é Benigno*).
 - **senão** (*saída é maligno*).

Para o desenvolvimento do classificador bayesiano, decidimos separar as 500 primeiras amostras da base de Wisconsin para servir de base de treinamento e realizar os cálculos probabilísticos. As demais 199 amostras foram utilizadas como amostras de teste.

Os classificadores foram desenvolvidos em linguagem Java.

Resultados

Os autores do artigo “Applying Fuzzy CoCo to Breast Cancer Diagnosis” [5] afirmaram ter alcançado uma taxa de acerto de 98,54% com a configuração fuzzy utilizada. O nosso classificador fuzzy conseguiu uma taxa próxima de 96,7% utilizando a mesma configuração. A diferença obtida pode ser devida ao fato de que, na época em que o trabalho original foi desenvolvido, a base de Wisconsin possuía 683 amostras diagnosticadas, enquanto atualmente possui 699 amostras. O acréscimo de 16 amostras pode ter impactado no resultado obtido, apesar da taxa de acerto ainda se mostrar alta.

Executando o Classificador Bayesiano com as 500 primeiras amostras direcionadas para o treinamento, obtemos uma taxa de acerto de mais de 99% ao classificar as 199 amostras restantes (apenas uma amostra classificada incorretamente). Ao variar a quantidade de amostras de treinamento de 300 a 500, obtemos uma pequena variação negativa na taxa de acerto, mas a mesma sempre se encontra acima dos 97%.

Conclusão

Os resultados obtidos com os classificadores são equiparáveis ou superiores aos melhores resultados encontrados na literatura para a base de dados de Wisconsin. A taxa de acerto do classificador fuzzy indica que os parâmetros fuzzy encontrados por Pena-Reyes e Sipper [5] realmente se adequam bem ao problema de Wisconsin, conseguindo prever de maneira ótima os diagnósticos positivos e negativos de aspirações de agulha fina, e oferecendo uma boa segunda opinião para especialistas da área de diagnósticos. Os resultados apresentados pelo classificador bayesiano simples para a base de Wisconsin são surpreendentemente muito bons, sugerindo que há um grande nível de independência entre os atributos

de massas de mama (um atributo não influencia diretamente no outro), ou mais provavelmente, que essa dependência não tem importância significativa no diagnóstico de massas malignas ou benignas.

Referências

- [1] Ball, J.E., T.W. Butler, and L.M. Bruce. Towards Automated Segmentation and Classification of Masses in Digital Mammograms. in International Conference of the IEEE EMBS. 2004. San Francisco, CA, USA.
- [2] Mavroforakis, M., et al., Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines. European Journal of Radiology, 2005. 54(1): p. 80-89.
- [3] Merz , C. J. and Murphy , P. M. UC Irvine Machine Learning Repository. 1987. Disponível em: <<http://archive.ics.uci.edu/ml/index.html>>. Acesso em 03 nov. 2012.
- [4] Mangasarian , O. L., Street , W. N., and Wolberg, W. H. Breast Cancer Wisconsin (Diagnostic) Data Set. 1995. Disponível em: <[http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))>. Acesso em 03 nov. 2012.
- [5] Pena-Reyes, C.A. and Sipper, M. Applying Fuzzy CoCo to breast cancer diagnosis. *Proceedings of the 2000 Congress on Evolutionary Computation CEC00 Cat No00TH8512 2*, 1168-1175. 2000.
- [6] D. Heckerman. Bayesian Networks for Knowledge Discovery. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy Editors. *Advances in Knowledge Discovery and Data Mining*, pages 273-305. MIT Press, 1996.
- [7] Pedro Domingos, Michael Pazzani. Beyond independence: Conditions for the optimality of the simple Bayesian classifier. *Proc. International Conference on Machine Learning*, 1996, pages 105-112, Morgan Kaufmann.
- [8] R. R. Yager and D. P. Filev. *Essentials of Fuzzy Modeling and Control*. John Wiley & Sons, Inc., 1994.
- [9] R. R. Yager and L. A. Zadeh. *Fuzzy Sets, Neural Networks, and Soft Computing*. Van Nostrand Reinhold, New York, 1994.