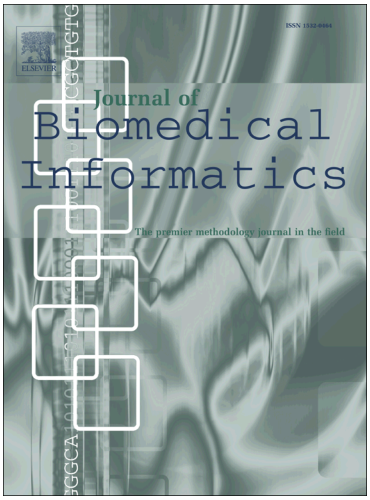

Gene selection for tumor classification using neighborhood rough sets and entropy measures

Yumin Chen, Zunjun Zhang, Jianzhong Zheng, Ying Ma, Yu Xue

PII: S1532-0464(17)30032-1
DOI: http://dx.doi.org/10.1016/j.jbi.2017.02.007
Reference: YJBIN 2725

To appear in: *Journal of Biomedical Informatics*

Please cite this article as: Chen, Y., Zhang, Z., Zheng, J., Ma, Y., Xue, Y., Gene selection for tumor classification using neighborhood rough sets and entropy measures, *Journal of Biomedical Informatics* (2017), doi: http://dx.doi.org/10.1016/j.jbi.2017.02.007

Gene selection for tumor classification using neighborhood rough sets and entropy measures

Yumin Chen [a], Zunjun Zhang [b,*], Jianzhong Zheng [b], Ying Ma [a], Yu Xue [c]

[a] *College of Computer & Information Engineering, Xiamen University of Technology, Xiamen 361024, China*

[b] *Department of Urinary Surgery, The Third Xiamen Hospital of Fujian University of Traditional Chinese Medicine, Xiamen 316000, China*

[c] *School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing 210044, China*

Abstract

With the development of bioinformatics, tumor classification from gene expression data becomes an important useful technology for cancer diagnosis. Since a gene expression data often contains thousands of genes and a small number of samples, gene selection from gene expression data becomes a key step for tumor classification. Attribute reduction of rough sets has been successfully applied to gene selection field, as it has the characters of data driving and requiring no additional information. However, traditional rough set method deals with discrete data only. As for the gene expression data containing real-value or noisy data, they are usually employed by a discrete preprocessing, which may result in poor classification accuracy. In this paper, we propose a novel gene selection method based on the neighborhood rough set model, which has the ability of dealing with real-value data whilst maintaining the original gene classification information. Moreover, this paper addresses