

2-4-2010

## A robust gene selection method for microarray-based cancer classification.

Xiaosheng Wang

University of Nebraska Medical Center, [xiaosheng.wang@unmc.edu](mailto:xiaosheng.wang@unmc.edu)

Osamu Gotoh

Kyoto University

Tell us how you used this information in this [short survey](#).

Follow this and additional works at: [https://digitalcommons.unmc.edu/com\\_gcba\\_articles](https://digitalcommons.unmc.edu/com_gcba_articles)



Part of the [Medical Anatomy Commons](#), [Medical Cell Biology Commons](#), and the [Medical Genetics Commons](#)

---

### Recommended Citation

Wang, Xiaosheng and Gotoh, Osamu, "A robust gene selection method for microarray-based cancer classification." (2010). *Journal Articles: Genetics, Cell Biology & Anatomy*. 11.

[https://digitalcommons.unmc.edu/com\\_gcba\\_articles/11](https://digitalcommons.unmc.edu/com_gcba_articles/11)

This Article is brought to you for free and open access by the Genetics, Cell Biology & Anatomy at DigitalCommons@UNMC. It has been accepted for inclusion in Journal Articles: Genetics, Cell Biology & Anatomy by an authorized administrator of DigitalCommons@UNMC. For more information, please contact [digitalcommons@unmc.edu](mailto:digitalcommons@unmc.edu).

**OPEN ACCESS**

Full open access to this and thousands of other papers at <http://www.la-press.com>.

## A Robust Gene Selection Method for Microarray-based Cancer Classification

Xiaosheng Wang<sup>1</sup> and Osamu Gotoh<sup>1,2</sup>

<sup>1</sup>Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto, Japan.

<sup>2</sup>National Institute of Advanced Industrial Science and Technology, Computational Biology Research Center, Tokyo, Japan.

Email: [david@genome.ist.i.kyoto-u.ac.jp](mailto:david@genome.ist.i.kyoto-u.ac.jp)

---

**Abstract:** Gene selection is of vital importance in molecular classification of cancer using high-dimensional gene expression data. Because of the distinct characteristics inherent to specific cancerous gene expression profiles, developing flexible and robust feature selection methods is extremely crucial. We investigated the properties of one feature selection approach proposed in our previous work, which was the generalization of the feature selection method based on the depended degree of attribute in rough sets. We compared the feature selection method with the established methods: the depended degree, chi-square, information gain, Relief-F and symmetric uncertainty, and analyzed its properties through a series of classification experiments. The results revealed that our method was superior to the canonical depended degree of attribute based method in robustness and applicability. Moreover, the method was comparable to the other four commonly used methods. More importantly, the method can exhibit the inherent classification difficulty with respect to different gene expression datasets, indicating the inherent biology of specific cancers.

**Keywords:** microarrays, cancer classification, feature selection, dependent degree, rough sets, machine learning

---

*Cancer Informatics* 2010:9 15–30

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



## Background

One major problem in applying gene expression profiles to cancer classification and prediction is that the number of features (genes) greatly surpasses the number of samples. Some studies have shown that a small collection of genes selected correctly can lead to good classification results.<sup>1-4</sup> Therefore gene selection is crucial in molecular classification of cancer. Numerous methods of selecting informative gene groups to conduct cancer classification have been proposed. Most of the methods first ranked the genes based on certain criteria, and then selected a small set of informative genes for classification from the top-ranked genes. The most used gene ranking approaches include t-score, chi-square, information entropy-based, Relief-F, symmetric uncertainty etc.

In,<sup>2</sup> we used a new feature selection method for gene selection. The feature selection method was based on the

ERROR: invalidaccess  
OFFENDING COMMAND: put

STACK:

/93

1

```
[/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/space /exclam /quotedbl /numbersign /dollar /percent /ampersand /  
quoteright /parenleft /parenright /asterisk /plus /comma /hyphen /period  
/slash /zero /one /two /three /four /five /six /seven /eight /nine /  
colon /semicolon /less /equal /greater /question /at /A /B /C /D /E /F /  
G /H /I /J /K /L /M /N /O /P /Q /R /S /T /U /V /W /X /Y /Z /bracketleft  
/backslash /bracketright /asciicircum /underscore /quoteleft /a /b /c /d  
/e /f /g /h /i /j /k /l /m /n /o /p /q /r /s /t /u /v /w /x /y /z /  
braceleft /bar /braceright /asciitilde /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
exclamdown /cent /sterling /fraction /yen /florin /section /currency /  
quotesingle /quotedblleft /guillemotleft /guilsinglleft /guilsinglright  
/fi /fl /.notdef /endash /dagger /daggerdbl /periodcentered /.notdef /  
paragraph /bullet /quotesinglbase /quotedblbase /quotedblright /  
guillemotright /ellipsis /perthousand /.notdef /questiondown /.notdef /  
grave /acute /circumflex /tilde /macron /breve /dotaccent /dieresis /  
.notdef /ring /cedilla /.notdef /hungarumlaut /ogonek /caron /emdash /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
AE /.notdef /ordfeminine /.notdef /.notdef /.notdef /.notdef /Lslash /  
Oslash /OE /ordmasculine /.notdef /.notdef /.notdef /.notdef /.notdef /  
ae /.notdef /.notdef /.notdef /dotlessi /.notdef /.notdef /lslash /  
oslash /oe /germandbls /.notdef /.notdef /.notdef /.notdef ]  
[/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef  
/space /exclam /quotedbl /numbersign /dollar /percent /ampersand /  
quoteright /parenleft /parenright /asterisk /plus /comma /hyphen /period  
/slash /zero /one /two /three /four /five /six /seven /eight /nine /  
colon /semicolon /less /equal /greater /question /at /A /B /C /D /E /F /  
G /H /I /J /K /L /M /N /O /P /Q /R /S /T /U /V /W /X /Y /Z /bracketleft  
/backslash /bracketright /asciicircum /underscore /quoteleft /a /b /c /d  
/e /f /g /h /i /j /k /l /m /n /o /p /q /r /s /t /u /v /w /x /y /z /  
braceleft /bar /braceright /asciitilde /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
exclamdown /cent /sterling /fraction /yen /florin /section /currency /  
quotesingle /quotedblleft /guillemotleft /guilsinglleft /guilsinglright  
/fi /fl /.notdef /endash /dagger /daggerdbl /periodcentered /.notdef /  
paragraph /bullet /quotesinglbase /quotedblbase /quotedblright /  
guillemotright /ellipsis /perthousand /.notdef /questiondown /.notdef /  
grave /acute /circumflex /tilde /macron /breve /dotaccent /dieresis /  
.notdef /ring /cedilla /.notdef /hungarumlaut /ogonek /caron /emdash /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /  
AE /.notdef /ordfeminine /.notdef /.notdef /.notdef /.notdef /Lslash /  
Oslash /OE /ordmasculine /.notdef /.notdef /.notdef /.notdef /.notdef /  
ae /.notdef /.notdef /.notdef /dotlessi /.notdef /.notdef /lslash /  
oslash /oe /germandbls /.notdef /.notdef /.notdef /.notdef ]
```