An Investigation into the use of Fourier Transform Infrared Spectroscopy for Breast Cancer Data Sets

By
Shabbar Naqvi
szn@cs.nott.ac.uk

Supervisor
Dr. Jonathan M. Garibaldi
jmg@cs.nott.ac.uk
Presentation Content

- Background
- Aims and objectives of current work
- Fourier Transform Infrared-Micro spectroscopy (FTIR)
- Difficulties
- Data set description
- Future work
Background

- Breast Cancer has the highest incidence rate in women around the world including U.K.

- Cancer prognosis

- Nottingham Prognostic Index (NPI)

  \[ \text{NPI} = (0.2 \times \text{tumour diameter in cm}) + \text{lymph node stage} + \text{tumour grade} \]

- NPI categorizes patients in three prognosis groups Good, Intermediate & Poor
Difficulties

- Complex and time consuming process
- High risk of human error.
- The best pathologists have been shown to exhibit variability
- Need to develop automated methods minimising the risks of false results
Aims & Objectives of Current Work

- To investigate the potential of Fourier Transform Infrared Micro-Spectroscopy (FTIR Microscopy) in providing an automated method for finding solutions with such problems.

- The objective of the current research is to develop a novel method with the help of FTIR microscopy and advanced computational methods of machine learning that has high performance, is reliable and is cost-effective, in order to be useful for clinicians in real clinical practice.
Fourier Transform Infrared Micro-Spectroscopy (FTIR Microscopy)

Motivation

- It has been increasingly applied to the study of biomedical conditions that includes differentiating between different types of cancer tissues.
- It is a very powerful tool for the determination and monitoring of chemical composition within biological systems.
An Example of FTIR absorbance Spectra
Data and Methods

Data

- New FTIR data sets were expected to be obtained from Department of Chemistry (Problems!!)

- New project with University of Illinois at Urbana-Champaign, USA with Prof. Rohit Bhargava

- Large data sets (Approximately 80GB per data set)
Problems

- How to transfer such large data sets? (Over 80GB)
- Pilot study by transferring one data set
- How to read the Data?
- Envi format?
- Try importing in Matlab or in R?
Data set Description

- Name: BR804 (size 85.4GB)
- 80 cores
- 40 cases
- Cancer grades break down
  - Grade1: 02
  - Grade2: 26
  - Grade3: 06
  - Not defined: 06

source: http://www.biomax.us/tissue-arrays/Breast/BR804
Data set Description

- Samples (x-axis) = 4062
- Lines (y-axis) = 3420
- Bands (z-axis) = 1641
- Each point is of type float (4 bytes)
- Total size = \textbf{85.4GB}
Breast tumor tissue microarray, containing 40 cases of paired breast invasive cancer and matched normal adjacent tissue, single core per case.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
<tr>
<td>B</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
<tr>
<td>C</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
<tr>
<td>D</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
<tr>
<td>E</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
<tr>
<td>F</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
<tr>
<td>G</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
<tr>
<td>H</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
<td>Bre</td>
</tr>
</tbody>
</table>

**Legend:**
- **Bre** - Breast
- **-** - Malignant tumor, **-** - NAT
Initial Success

- Development of R-script
- At last we can see some real FTIR spectra
FTIR Spectra

NAT(M)  Malignant (M)
Pre-processing of FTIR Spectra

- Base line correction
- Normalisation
Normalisation

FTIR Spectra at 10 selected points

School of Computer Science
Faculty of Science
Normalisation

Normalized Spectra

Absorbance

Wavenumbers

School of Computer Science
Faculty of Science
Conclusion

- Background of the importance of breast cancer prognosis and difficulties
- The feasibility of using FTIR microscopy to develop an automated method
- Description of new data set and difficulties involved
Future Work

- Collaboration with the *Breast Cancer Pathology Research Group* of the University of Nottingham to identify points of interest
- Pre-processing of FTIR spectra of selected points
- Dimensionality reduction techniques (PCA)
- Use of machine learning methods
- Focus may be on SSL
- Application of results with NPI
Questions and Comments?
Thank you