Usage guidelines

Please refer to the usage guidelines at http://sure.sunderland.ac.uk/policies.html or alternatively

contact sure@sunderland.ac.uk.

University of Sunderland

"Performing Trend Analysis On Spatio-temporal Proteomics Data Using Differential Ratio Data Mining"

James Malone, Ken McGarry, Chris Bowerman
School of Computing and Technology
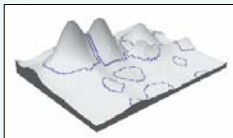
2-DE Gel Analysis

- Two-dimensional electrophoresis (2-DE) is unrivalled as a technique to analyse protein expression and is a key component of current proteomics research.
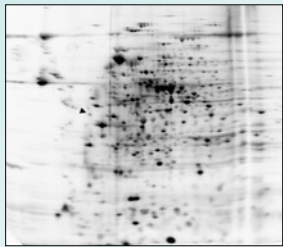
- A single gel (left) contains large amounts (~1,500) of spatial protein spot data (black dots).

- An experiment is often conducted over time, using several gels at each time point, creating large amounts of **spatio-temporal** data.

- Current analysis of such data is normally conducted manually which is

both time consuming and requires considerable expertise.

Differential Ratio Data Mining

- Differential ratio (dFr) data mining is a technique we have developed to perform an analysis on spatio-temporal datasets (of which 2-DE gels are one example).

- The dFr equation (right) is used to calculate a single measure, where; x and y are item variables, t is time point and dFr is the differential ratio produced at time point t.

$$\log\left\{\frac{\left(\frac{x_t}{y_t}\right)}{\left(\frac{x_{t+1}}{y_{t+1}}\right)}\right\} = dFr_t$$

- The overall data mining process is outlined below; the images are first transformed into data and then the algorithm is applied to the entire dataset over the time series. Finally, graphical representations are produced to assist the users interpretation.
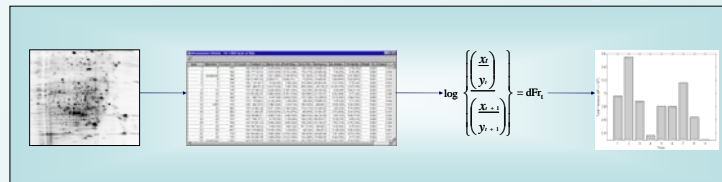

Experimental Results

- Two sets of experiments were performed on the 2-DE data; (i) to find significantly variant items over time and (ii) to determine significant time points.
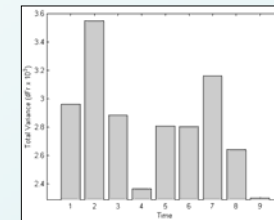
- Results of experiment (i) are shown left. This indicated that time points 2 and 7 were when most variation occurred.

- Results of experiment (ii) are shown right. This identified several proteins with high variance. These were also analysed using traditional

manual techniques which confirmed the findings to be valid.
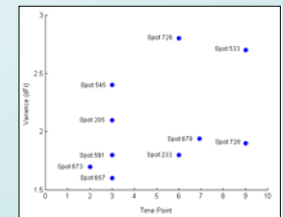
- This important piece of knowledge extraction highlights interesting proteins for further analysis, eliminating the need to analyse thousands of uninteresting proteins, thereby reducing time and cost.