East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.280909.270410a

Vol. **1**, No. 1, pp. 35-48
February 2011

A Comparative Method for Analysing Toponome Image Stacks

Andrei Barysenka[1,*], Andreas W. M. Dress[1] and Walter Schubert[2]

[1] *Department of Combinatorics and Geometry, CAS-MPG Partner Institute and Key Lab for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China and Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany.*
[2] *Molecular Pattern Recognition Research Group Medical Faculty, Otto-von-Guericke University, Leipziger Str. 44 Magdeburg, 39120 Germany.*

Received 28 September 2009; Accepted (in revised version) 27 April 2010

Available online 26 October 2010

Abstract. We present a technique to find threshold values that allows the user to separate signal from noise in fluorescence grey-level images. It can be classified as a purely *comparative* method based upon the amount of "Mutual Information" between two or more florescence images, and we apply it to stacks of such images produced using the newly-developed MELK technology. Our results are compared to results obtained by another research group using a quite different (completely independent and more technology-based) approach; and also to results obtained using *Otsu's Thresholding Method*, yet another completely distinct approach invented to separate foreground and background in a grey-level image, based on minimising "intra-class variance" [9, 10]. The remarkably good agreement found suggests that our proposed comparative information based method not only accounts for the biological mechanisms governing cellular protein networks very well, but also (and probably much more importantly) shows that cells actually organise the spatial structure of their protein networks in a highly non-random fashion as might be expected – and thereby try to optimise their "mutual information content", and thus most probably their efficiency.

Key words: MELK technology, Kullback-Leibler distance, threshold values, numerical optimization.

1. Introduction

Given a finite collection $\mathscr{C}$ of observations, each observation $c \in \mathscr{C}$ being represented by a vector $\varphi(c) = (c_1, \cdots, c_n) \in \mathbb{R}^n$, there are many areas – from sensing, security, and data mining to biology and medicine – where a pre-processing step assigning a 0, 1-vector

*Corresponding author. *Email addresses:* aboris@picb.ac.cn (A. Barysenka), andreas@picb.ac.cn (A. W. M. Dress), walter.schubert@med.ovgu.de (W. Schubert)

$\boldsymbol{\chi}(c) = \bar{c} = (\bar{c}_1, \cdots, \bar{c}_n)$ to each observation $c \in \mathscr{C}$ can be a great advantage. Clearly, it is appropriate that this assignment be *monotone* (i.e. $c, d \in \mathscr{C}$, $i \in \{1, \cdots, n\}$), that $c_i \leq d_i$ should always imply $\bar{c}_i \leq \bar{d}_i$, and that it should maximise the "surprise" (or the "mutual information content") that can be found in the data, following the "surprisology" paradigm proposed in [14].

The monotonicity requirement is easily met, by choosing a threshold or 'cut-off' vector $\mathbf{t} = (t_1, \cdots, t_n) \in \mathbb{R}^n$ that allows us to define a monotone map

$$\boldsymbol{\psi}_{\mathbf{t}} : \mathbb{R}^n \to \{0,1\}^n : (x_1, \cdots, x_n) \mapsto (\bar{x}_1, \cdots, \bar{x}_n)$$

by putting

$$\bar{x}_i := \begin{cases} 1 & \text{if } x_i \geq t_i, \\ 0 & \text{otherwise} \end{cases}$$

for all $i = 1, \cdots, n$. When composed with $\boldsymbol{\varphi}$, this yields the required $\{0,1\}$ vector $\bar{\mathbf{c}} = \boldsymbol{\psi}_{\mathbf{t}}(c_1, \cdots, c_n) = \boldsymbol{\psi}_{\mathbf{t}}\big(\boldsymbol{\varphi}(c)\big)$ for each observation $c \in \mathscr{C}$.

We follow the basic exposition already outlined in [2], by assuming that the mutual independence of the observed values $c_i, i \in \{1, \cdots, n\}$ of any observation $c \in \mathscr{C}$ (i.e. over all observations in $\mathscr{C}$) would *not* yield a surprise. Then one can measure the 'amount of surprise' encountered when associating a $\{0,1\}$ vector $\boldsymbol{\chi}(c) = \big(\chi_1(c), \cdots, \chi_n(c)\big)$ to each $c \in \mathscr{C}$, by comparing the *observed* probability distribution p_χ defined for each $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n) \in \{0,1\}^n$ by

$$p_\chi(\boldsymbol{\varepsilon}) := \frac{\#\{c \in \mathscr{C} | \boldsymbol{\chi}(c) = \boldsymbol{\varepsilon}\}}{\#\mathscr{C}} \tag{1.1}$$

with the *expected* probability distribution q_χ defined by

$$q_\chi(\boldsymbol{\varepsilon}) := \prod_{i=1}^{n} \frac{\#\{c \in \mathscr{C} | \chi_i(c) = \varepsilon_i\}}{\#\mathscr{C}}. \tag{1.2}$$

Given the independence of the coordinates, note that q_χ would essentially coincide with p_χ. Consequently, noting that $\boldsymbol{\varepsilon} \in \{0,1\}^n$ and $q_\chi(\boldsymbol{\varepsilon}) = 0$ always implies $p_\chi(\boldsymbol{\varepsilon}) = 0$, the well-known *Kullback-Leibler Divergence* or *Mutual-Information Function*

$$MI(q_\chi \to p_\chi) := \sum_{\boldsymbol{\varepsilon} \in \{0,1\}^n} p_\chi(\boldsymbol{\varepsilon}) \log\left(\frac{p_\chi(\boldsymbol{\varepsilon})}{q_\chi(\boldsymbol{\varepsilon})}\right) \tag{1.3}$$

(with $p_\chi(\boldsymbol{\varepsilon}) \ln(p_\chi(\boldsymbol{\varepsilon})/q_\chi(\boldsymbol{\varepsilon})) := 0$ for $p_\chi(\boldsymbol{\varepsilon}) = 0$) appears to be a good measure of that 'amount of surprise' encountered by the map $\boldsymbol{\chi} : \mathscr{C} \to \{0,1\}^n$ (see e.g., [11], for its various virtues) . Thus to each cut-off vector $\mathbf{t} \in \mathbb{R}^n$ one can associate its surprise value $surp(\mathbf{t}) = surp_{(\mathscr{C}, \varphi)}(\mathbf{t})$ defined by

$$\mathrm{surp}(\mathbf{t}) := MI(q_{\psi_{\mathbf{t}} \circ \varphi} \to p_{\psi_{\mathbf{t}} \circ \varphi}), \tag{1.4}$$

and declare a cut-off vector $\mathbf{t}$ to be $(\mathscr{C}, \boldsymbol{\varphi})$-optimal if

$$\mathrm{surp}(\mathbf{t}) \geq \mathrm{surp}(\mathbf{t}')$$