3074
doi:10.1017/S143192762201145X
Microsc. Microanal. 28 (Suppl 1), 2022

Data Analytics: Quality Measures for Image Information Content

Carol A. Heckman[1]

[1.] Center for Microscopy and Microanalysis, Bowling Green State University, Bowling Green, OH USA.

Methods of evaluating images are being developed in many specialized areas of science but are particularly important in microscopy. Here, the data may consist largely or exclusively of images. In parallel with the methods for evaluating objects, measures of the quality of the information are needed. For large data assemblages in other fields, initial procedures to reduce dimensionality may be applied. This reduction ensures that the number of objects in the sample vastly exceed the number of variables used for the data analysis. The purpose of this work was to determine whether dimensional reduction is useful for making a statistical determination that can serve as a measure of information quality.

Multiple replicates of an experiment were done. Data from the experimental and control groups were compared on a statistical basis. A sample consisted of silhouettes of 50 cells per treatment group. The cells' outlines, derived by exact segmentation, were used to extract information, for example, variables for shape. The latent factors were derived by reduction [1-3], as illustrated in Fig. 1. Three variables, factors 4, 5, and 7, were chosen for comparison, and they were ranked in order of their importance. A phenotype of the sample was assigned based on the combinations of these factors. As shown in Fig. 2, each treated sample must fall into one of 27 possible phenotypes.

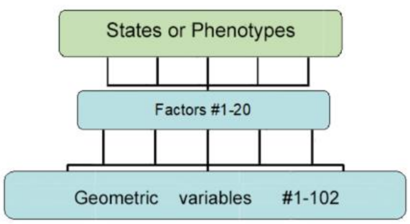

Figure 1. Method of mapping variables to phenotypes.

All the experiments showed a unique pattern of change in the phenotype due to the treatment (Figs. 3-5). In each experiment, however, at least one factor was changed significantly compared to the control. This finding suggests a robust test for differences in an experiment. If the coarse-grained variables obtained by extracting principal components or latent factors are tested first, they can demonstrate that a change was introduced by the treatment.

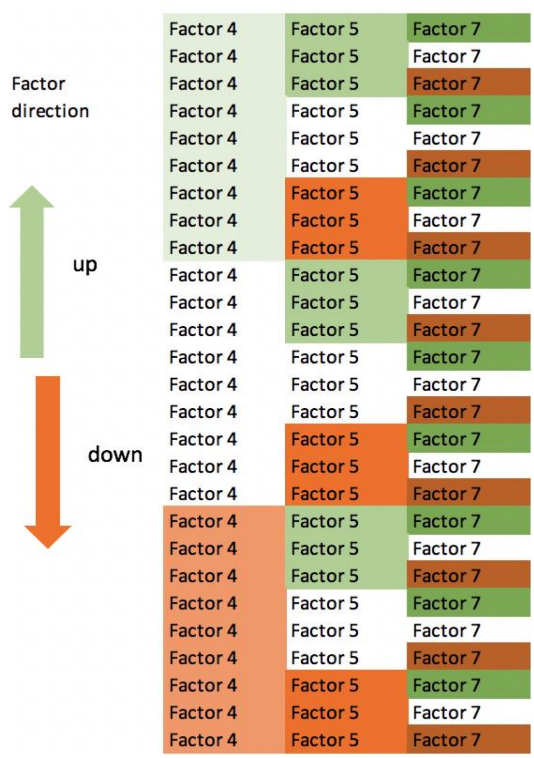

Figure 2. All possible combinations of Factors 4, 5, and 7. Green colors indicate that a value is significantly increased compared to control. Orange or brown colors indicate a significant decrease.

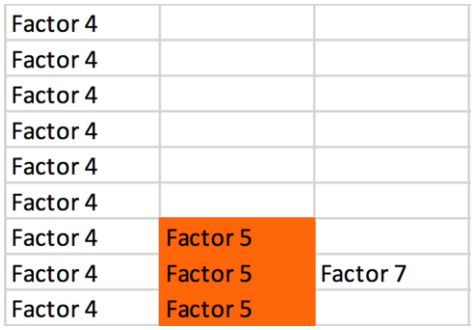

Figure 3. Reduced dimensionality variables from a treatment with Taxol and colchicine over control. In experiment #1, the Factor 5 values were decreased in the treated sample.

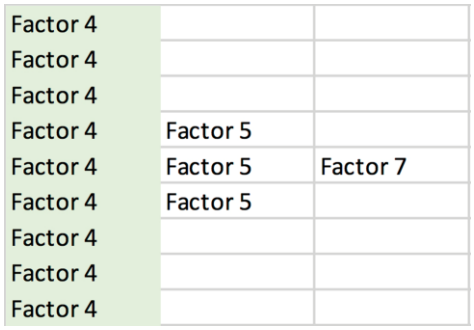

Figure 4. In experiment #2, the Factor 4 values were significantly higher in the treated sample.

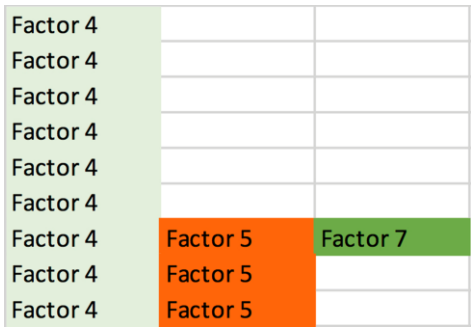

Figure 5. In experiment #3, both changes noted in the previous experiments occurred along with an additional elevation of Factor 7

When comparing samples, biologists typically select individual cells or organisms with a view to ensuring that the number is high enough to measure a statistic. Automated methods of data acquisition allow thousands of individuals to be sampled, while the number of variables can exceed a hundred. With cultured cells, this design almost always reveals a difference between experimentals and controls. To assess variations in assay conditions and/or sample preparation requires a greater effort.

By using machine learning, biologists are able to automatically and routinely detect objects in images and derive measurements from them. As there is a dearth of quality standards for image data, it has become important to investigate ways of assessing information content. For example, how do investigators know that the information content is sufficient to distinguish among treatment groups? And how do they know how many images must be analyzed to arrive at a statistically valid conclusion? These problems are only exacerbated in multidimensional modes of imaging. These findings may have applications in fields other than biology, in fact, wherever there is a need for rigorous quality measures.

References:

[1] CA Heckman and RJ Jamasbi, Experimental Cell Research **246** (1999), p. 69.
[2] CA Heckman et al., Cellular Signalling **24** (2012), p. 189.
[3] M Varghese et al., International Journal Cell Biology [Online] **2012,** Article ID 402916 (2012), p. 16. doi: 10.1155/2012/402916

https://doi.org/10.1017/S143192762201145X Published online by Cambridge University Press