INVERSION AND INFORMATICS COMBINED: MAXIMISING BENEFIT FROM GEO- EXPERIMENT THROUGH COMPUTATION

A.M. Reading & M.J. Cracknell

School of Physical Sciences (Earth Sciences), Private Bag 79, University of Tasmania, Hobart, TAS 7001, Australia

Key words inversion, informatics, high-dimensional, solution space.

Many of the great and diverse challenges of understanding the natural, physical world use computational approaches to match a model to experimental data, through a process of geophysical inversion [1, 2]. Solving inverse problems has proven successful in geophysics with recent developments including thorough investigations of the solution space and uncertainty estimates [3]. This contribution explores complimentary, additional, approaches to learning from data through informatics approaches familiar to the information and communications technology (ICT) community. Such approaches work well with large volume and/or high-dimensional datasets, and include the possibility of combining categorical and numerical information. We show examples of how combining inversion and informatics, in computational geophysics, enables the benefit from experiments and resulting datasets to be maximised.

While approaches to solving inverse problems are familiar in the computational geophysics community, informatics approaches are less widely used. Learning from data in this frame may be accomplished through supervised or unsupervised learning, which corresponds respectively to a deductive or inductive underlying philosophy. In the former case, the predictive algorithm is developed through the use of a training dataset. Of the multiple algorithms [4] tested for this purpose, Random Forests [5] has proven a good first choice for geoscience datasets. Of the unsupervised algorithms, self-organising maps [6] has seen increasing use in the natural physical sciences. Key to the transfer of informatics approaches from ICT to geophysics is a robust process of prediction evaluation.

Combining inversion and informatics approaches is particularly valuable where data fusion across high-dimensional datasets is desired. This enables information from mixed-category data to be combined, thus making best use of data from field campaigns and/or remote sensing datasets. Using combined approaches provides not only a more complete approach to learning from hard earned data, but also a means of planning the next generation of field deployments or other experiments for maximum benefit. As we move from reconnaissance studies to the framing of first- and second-order experimental campaigns addressing particular hypotheses, the need to maximise the value of existing information and optimise ongoing data collection has never been greater.

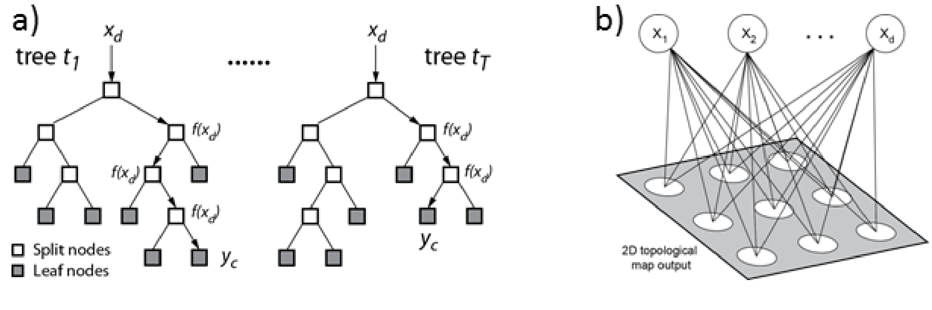

Figure 1. Representations of two informatics algorithms [5, 6] with potential wide application to geoscience data: a) Random Forests and b) self-organising maps.

References

[1] Aster, R.C., Borcher, B., and Thurber, C.H., 2013. *Parameter estimation and inverse problems*. – 2nd ed., Academic Press.

[2] Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*, Soceity of Industrial and Applied Mathematics.

[3] Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P., 2012. Transdimensional tomography with unknown data noise. *Geophysical Journal International*, v. 189, 1536-1556.

[4] Cracknell, M.J., and Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers and Geosciences*, v. 63, 22-33.

[5] Breiman, L., 2001. Random Forests. *Machine Learning*, v. 45, 5-32.

[6] Kohonen, T., 2001. *Self-organising maps*, Springer-Verlag.