Neural network hydrological modelling: on questions of over-fitting, over-training and over-parameterisation

R.J. ABRAHART (1), C.W. DAWSON (2), A.J. HEPPENSTALL (3), and L.M. SEE (3)

(1) School of Geography, University of Nottingham, Nottingham, United Kingdom (bob.abrahart@nottingham.ac.uk), (2) Department of Computer Science, Loughborough University, Loughborough, United Kingdom, (3) School of Geography, University of Leeds, Leeds, United Kingdom

The most critical issue in developing a neural network model is generalisation: how well will the preferred solution perform when it is applied to unseen datasets? The reported experiments used far-reaching sequences of model architectures and training periods to investigate the potential damage that could result from the impact of several interrelated items: (i) over-fitting - a machine learning concept related to exceeding some optimal architectural size; (ii) over-training - a machine learning concept related to the amount of adjustment that is applied to a specific model - based on the understanding that too much fine-tuning might result in a model that had accommodated random aspects of its training dataset - items that had no causal relationship to the target function; and (iii) over-parameterisation - a statistical modelling concept that is used to restrict the number of parameters in a model so as to match the information content of its calibration dataset. The last item in this triplet stems from an understanding that excessive computational complexities might permit an absurd and false solution to be fitted to the available material. Numerous feedforward multilayered perceptrons were trialled and tested. Two different methods of model construction were also compared and contrasted: (i) traditional Backpropagation of Error; and (ii) state-of-the-art Symbiotic Adaptive Neuro-Evolution. Modelling solutions were developed using the reported experimental set ups of Gaume & Gosset (2003). The models were applied to a near-linear hydrological modelling scenario in which past upstream and past downstream discharge records were used to forecast current discharge at the downstream gauging station [CS1: River Marne]; and a non-linear hydrological modelling scenario in which past river discharge measurements and past local meteorological records (precipitation and evaporation) were used to forecast current discharge at the river gauging station [CS2: Le Sauzay].