Title

Application of artificial intelligence techniques for the determination of groundwater level using spatio-temporal parameters

Authors

Amirhossein Najafabadipour [*1]. Gholamreza Kamali [1]. Hossein Nezamabadi-pour [2]

Authors' Affiliations

[1] Department of Mining Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

[2] Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

Corresponding author

Amirhossein Najafabadipour, Department of Mining Engineering, Shahid Bahonar University of Kerman, Kerman, Iran. Email: najafabadipour@eng.uk.ac.ir

Optimization Techniques

Levenberg-Marquardt Algorithm (LM)

The most common tool used to optimize the weight and bias of multilayer perceptron and cascade forward is the Levenberg-Marquardt (LM) algorithm, also known as the damped least-squares method [1]. This algorithm is used to solve nonlinear least-squares problems that find local minimums. This method does not require to calculate of the Hessian matrix and the gradient is calculated from the following equation [2]:

$$H = J^T J \tag{1}$$

$$g = J^T e \tag{2}$$

where e stands a vector of network errors, and J expresses a Jacobian matrix. In the following relation of updating the LM algorithm, the mentioned approximation with the Hessian matrix is used:

$$x_{k+1} = x_k - (J^T J - \eta I)^{-1} J^T e \tag{3}$$

η is a constant, and x denotes connection weights. η increases when an experimental step enlarges the efficiency function.

Bayesian Regularization Algorithm (BR)

The Bayesian Regularization (BR) training algorithm, according to Levenberg-Marquardt optimization, updates weights and biases by minimizing a combination of squared errors and weights [3]. Afterward, BR calculates the right combination to develop a network with superior generalization [4]. Network weights are expressed as a training cost function by the BR algorithm using the following equation:

$$F(\omega) = \alpha E_\omega + \beta E_D \tag{4}$$

In which E_D and E_ω are the sum of the network errors and the sum of the squared network weights, respectively, $F(\omega)$ denotes the objective function. In the BR optimizer, the network weights are random variables in which the network weights and the training sets have a Gaussian distribution. α and β Factors are objective function parameters that are clarified based on Bayes' theorem.

Scaled Conjugate Gradient Algorithm (SCG)

One of the basic features of the backpropagation algorithm is to reach the most negative gradient, and it uses the adjustment of weights in the steepest descending direction [5]. Along such a direction a decrease in function performance is observed faster but does not cause faster convergence. In this direction, a search for the conjugate gradient (CG) method leads to faster convergence than the steepest descending direction and the error minimization is maintained in the previous steps [6].

$$P_0 = -g_0 \tag{5}$$

P is search direction, and $-g_0$ denotes the steepest descent direction in the first iteration. This direction is called conjugate direction, commonly used by conjugate gradient algorithms with the search line [7]. To evaluate the optimal distance to move in the current search direction, the step size is determined by a line search technique [6], which is shown by the following equation:

$$x_{k+1} = x_k + \alpha_k g_k \tag{6}$$

In other words, the proper search direction is calculated in a way that conjugates with the previous search direction.

$$P_k = -g_k + \beta_k P_{k-1} \tag{7}$$

Different versions of the conjugate algorithm are distinguished in the way that β is calculated [8].

Resilient Backpropagation Algorithm (RB)

The most widely used transfer functions in multilayer perceptron neural networks are Sigmoid and Tansig, which compress an infinite input range into a finite output. When using the steepest descent to train the network using these activation functions, the slope is small when an extensive input enters the function, leading to slight changes in weights and biases. The Resilient backpropagation method is used to remove the adverse effect of the partial derivatives, which is specified by the derivatives only for the direction of updating weights [9].

References

(1) Hagan, M. T.; Menhaj, M. B. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks* **1994**, *5* (6), 989-993. DOI: 10.1109/72.329697.

(2) Sahoo, S.; Jha, M. K. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. *Hydrogeol. J.* **2013**, *21* (8), 1865-1887. DOI: 10.1007/s10040-013-1029-5.

(3) MacKay, D. J. Bayesian interpolation. *Neural comput.* **1992**, *4* (3), 415-447.

(4) Foresee, F. D.; Hagan, M. T. Gauss-Newton approximation to Bayesian learning. In *Proceedings of International Conference on Neural Networks (ICNN'97)*, 1997; IEEE: Vol. 3, pp 1930-1935.

(5) Rostami, A.; Hemmati-Sarapardeh, A.; Shamshirband, S. Rigorous prognostication of natural gas viscosity: Smart modeling and comparative study. *Fuel* **2018**, *222*, 766-778. DOI: https://doi.org/10.1016/j.fuel.2018.02.069.

(6) Kisi, O.; Uncuoğlu, E. Comparison of three back-propagation training algorithms for two case studies. *Indian J. Eng. Mater. Sci.* **2005**, *12*.

(7) Møller, M. F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* **1993**, *6* (4), 525-533. DOI: https://doi.org/10.1016/S0893-6080(05)80056-5.

(8) Azadi, S.; Amiri, H.; Ataei, P.; Javadpour, S. Optimal design of groundwater monitoring networks using gamma test theory. *Hydrogeol. J.* **2020**, *28* (4), 1389-1402. DOI: 10.1007/s10040-020-02115-z.

(9) Riedmiller, M. A.; Braun, H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. *IEEE International Conference on Neural Networks* **1993**, 586-591 vol.581.