Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)

Optimizing the neural network hyperparameters utilizing genetic algorithm

Saeid Nikbakht
Cosmin Anitescu
Timon Rabczuk

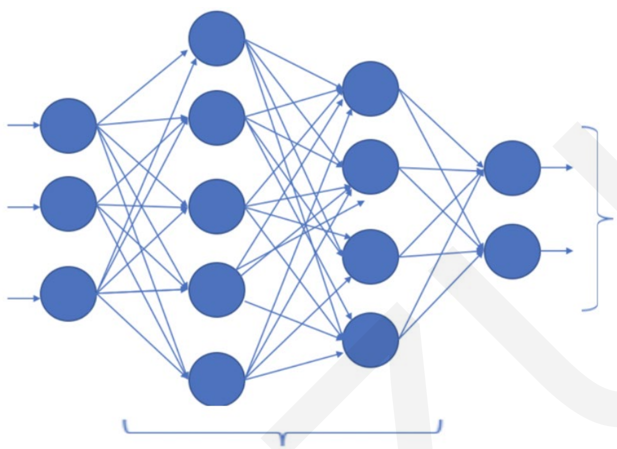
Cite this as: Saeid Nikbakht, Cosmin Anitescu, Timon Rabczuk, 2021. Optimizing the neural network hyperparameters utilizing genetic algorithm. *Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)*, 22(6):407-426.
https://doi.org/10.1631/jzus.A2000384

Deep Neural Networks (DNN)

Feed Forward

A successive matrix multiplication between network parameters and input is conducted to achieve the loss function

Input



Predicted Output

Hidden Layers containing network parameters

Loss Function :

The difference between true output and predicted output

Back-propagation

The network parameters are modified during the process of back-propagation with the aid of an optimizer in order to minimize the loss function

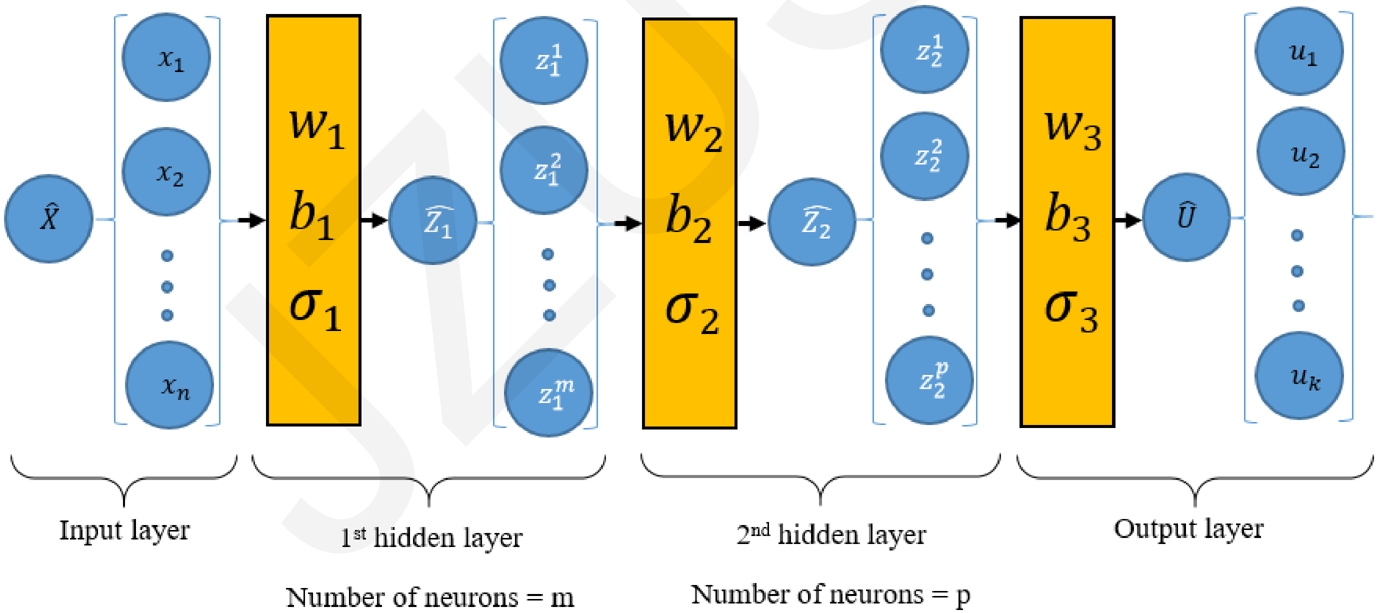
Hyper-Parameters VS Parameters

Network Parameters : W_i , b_i

- Chosen randomly at the beginning
- Trained through Back-propagation

Network Hyper-parameters :

- Number of hidden layers
- Number of Neurons in each layer [N1,N2,N3]
- Number of integration points {T1,T2,T3}
- Activation function (σ_i)
- Optimizer



Deep Energy Method (DEM)

Implementation of DNN for solving a partial differential equation considering boundary conditions

- The principle of minimum potential energy (acting as the partial differential equation):

$$\varepsilon = W_{\text{int}} - W_{\text{ext}},$$

$$W_{\text{int}} = \frac{1}{2}\int_{\Omega} \epsilon(u) : \boldsymbol{C} : \epsilon(u)\, d\Omega,$$

$$W_{\text{ext}} = \int_{\Omega} f \cdot u \, d\Omega + \int_{\partial\Omega_{\text{N}}} t_{\text{N}} \cdot u \, d\Gamma$$

- Relative strain energy error:

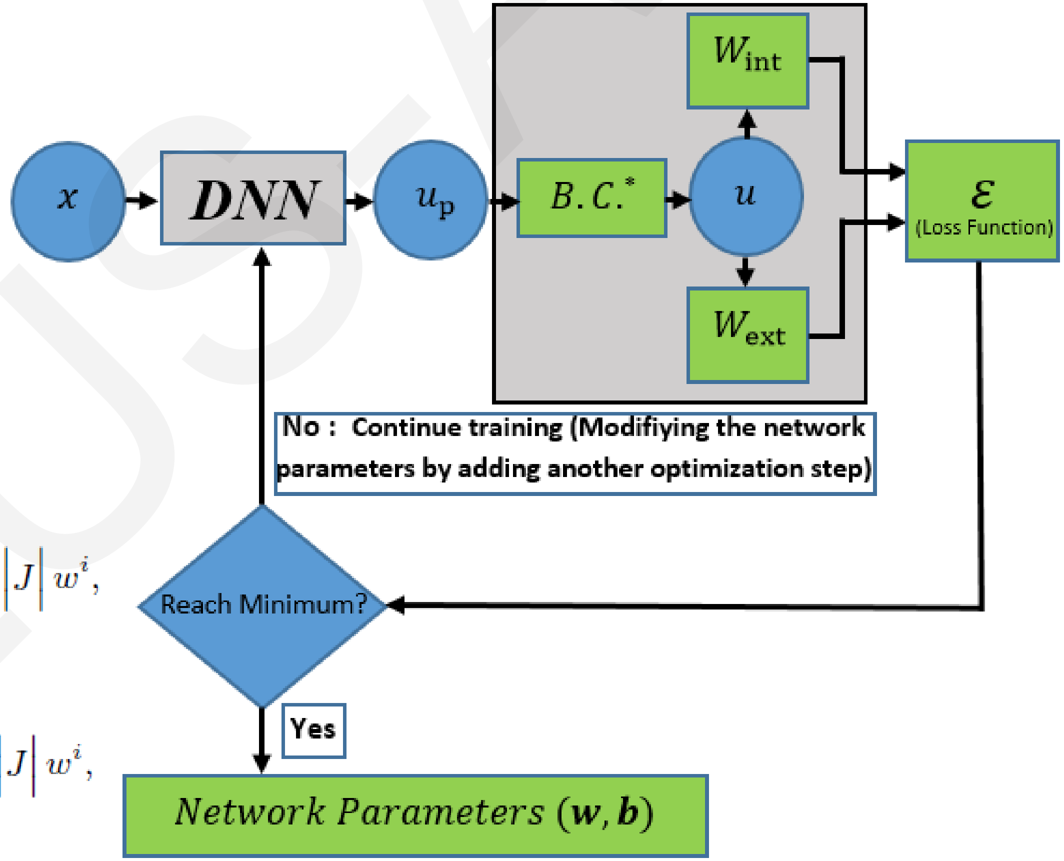
$$L_2^{\text{rel}}(W_{\text{int}}) = \sqrt{\frac{E_{\text{error}}}{E_{\text{norm}}}},$$

$$E_{\text{norm}} = \sum_{i=1}^{N_{\text{pred}}} \left[\begin{bmatrix} \sigma_{xx}^i \\ \sigma_{yy}^i \\ \sigma_{xy}^i \end{bmatrix}_{\text{exact}} [C]^{-1} \begin{bmatrix} \sigma_{xx}^i \\ \sigma_{yy}^i \\ \sigma_{xy}^i \end{bmatrix}_{\text{exact}}^{\text{T}} \right] |J|\, w^i,$$

$$E_{\text{error}} = \sum_{i=1}^{N_{\text{pred}}} \left[\begin{bmatrix} \sigma_{xx}^i \\ \sigma_{yy}^i \\ \sigma_{xy}^i \end{bmatrix}_{\text{error}} [C]^{-1} \begin{bmatrix} \sigma_{xx}^i \\ \sigma_{yy}^i \\ \sigma_{xy}^i \end{bmatrix}_{\text{error}}^{\text{T}} \right] |J|\, w^i,$$

- The Boundary Conditions:

$$u = \bar{u} \quad on\ \partial\Omega_{\text{D}}, \quad \sigma \cdot N = t_{\text{N}} \quad on\ \partial\Omega_{\text{N}},$$

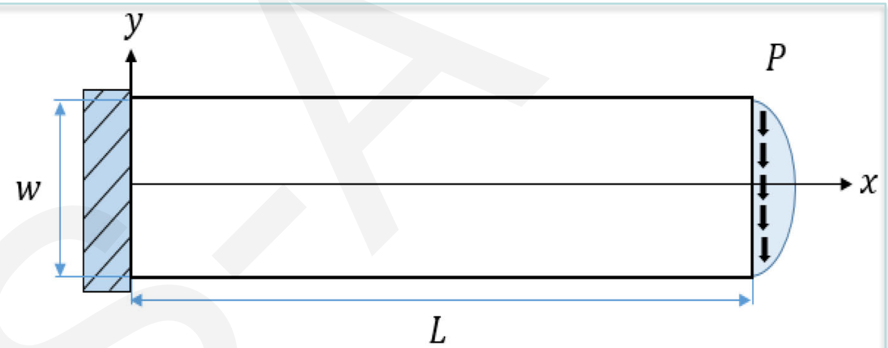


Deep Energy Method (DEM)

Case studies :

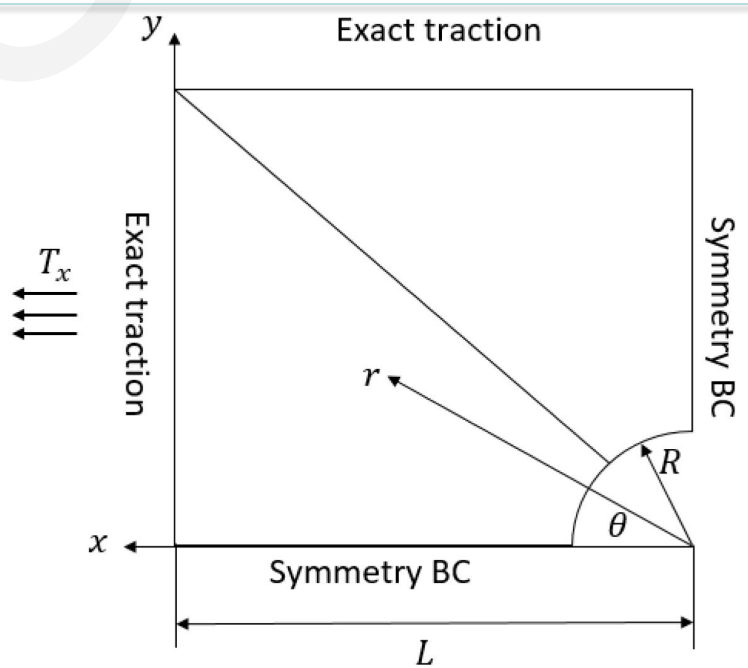
- Timoshenko cantilever beam :

This structure is modeled using NURBS to generate integration points in the x and y directions and on the boundary {T1,T2,T3}



- Plate with a hole in the middle :

This structure is modeled using NURBS to generate integration points in the radial and angular directions and on the boundary {T1,T2,T3}

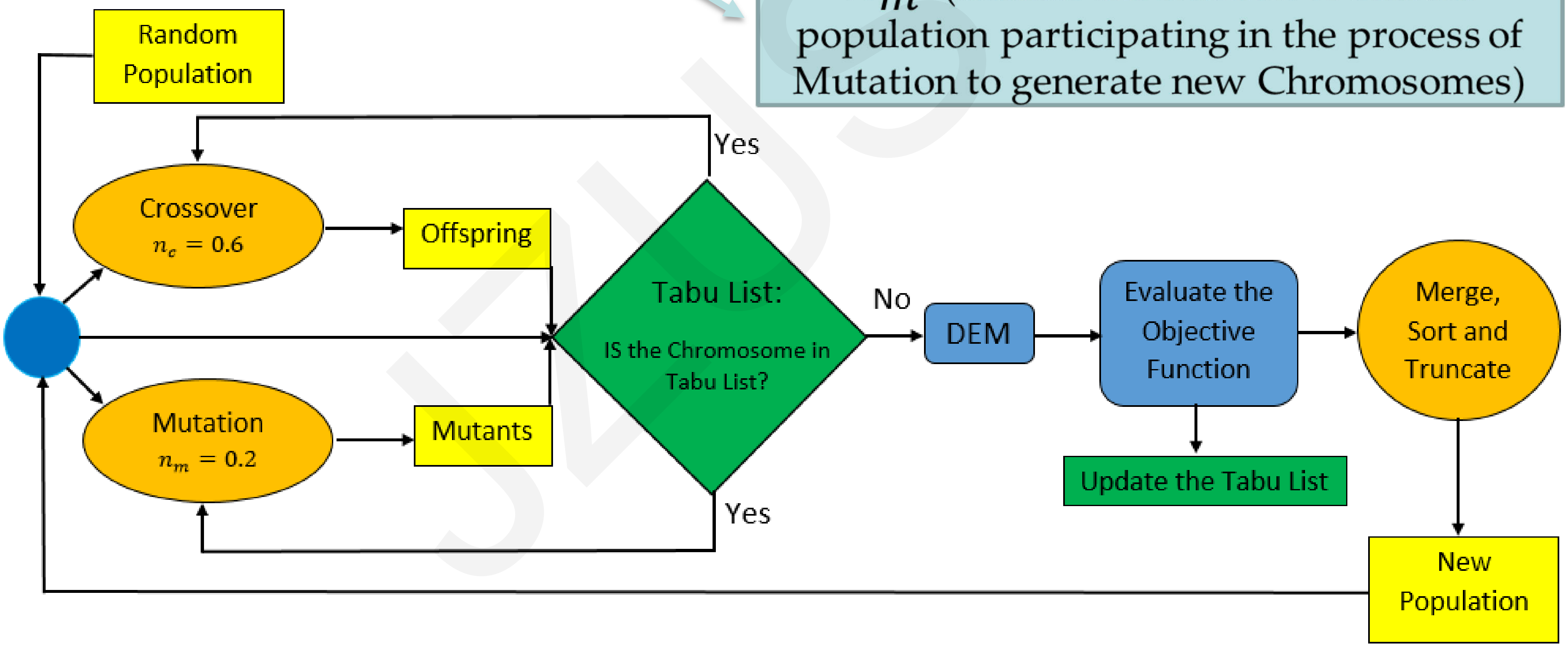


Genetic Algorithm (GA)

Genetic Algorithm parameters

n_c (defines the fraction of current population participating in the process of Crossover to generate new Chromosomes)

n_m (defines the fraction of current population participating in the process of Mutation to generate new Chromosomes)



Optimizing Hyper-Parameters

Activation functions :

- Sigmoid
- tanh
- ReLU
- $ReLU^2$ ✓

Optimizers :

- RMS
- Adam
- Adam + L-BFGS-B ✓

DEM optimum hyper-parameters:

- Timoshenko cantilever beam:

Network Hyper-parameters	Integration points	$L_2^{rel}(W_{int})$
[30,30,30]	{160,40,80}	0.0281
[41,39]	{160,40,80}	0.0131
[46,24,43]	{160,40,80}	0.0173
[30,19,17,17]	{160,40,80}	0.016

- Plate with a central hole:

Network Hyper-parameters	Integration points	$L_2^{rel}(W_{int})$
[30,30,30]	{160,40,80}	0.0281
[40,40,40]	{84,84,76}	0.0138
[30,41,44]	{80,80,80}	0.0151
[35,43,45]	{70,70,70}	0.0165

Conclusions

- DEM method is a robust and accurate method for solving partial differential equations
- Among activation functions, $ReLU^2$ leads to the most accurate results
- Among optimizers, Adam combined with L-BFGS-B leads to the most accurate results
- Both integration points and number of neurons in each layer have significant influence on the accuracy of DEM model, when calculating the stress distribution through structures
- Optimizing the hyper-parameters of DEM could lead to more than 50% and 38% decrease in the relative energy error of Timoshenko beam and plate with the central hole, respectively.
- The accuracy of predicting the stress and displacement distribution through the analyzed structures increases by decreasing the relative error of strain energy
- Although the process of finding the optimum hyper-parameters in DNNs has large computational costs, the remarkable influence of the optimized hyper-parameters on the accuracy of model could be considered as a valuable compensation.