Годишник на секция "Информатика"
Съюз на учените в България
Том 6, 2013, 39–45

Annual of "Informatics" Section
Union of Scientists in Bulgaria
Volume 6, 2013, 39–45

MODELING VERIFICATION PROCESS IN NEURAL NETWORK MULTILAYER PERCEPTRON WITH GENERALIZED NETS

Stanimir Surchev, Sotir Sotirov

Prof. Asen Zlatarov University, Burgas, Bulgaria
e-mails: ssurchev@gmail.com, ssotirov@btu.bg

Abstract: The proposed generalized net model presents a verification process in Multilayer Neural Networks. The purpose of the verification is to protect the neural network from overfitting. This process is one the most used in neural network training. There are many methods of verification and one of them is Early Stopping. It has a function to stop training when the network begins to overfit too often.
Keywords: Neural network, Generalized nets, Supervise learning, Overfitting, Early Stopping

1. Introduction

Neural Networks is an abstract represent of human brain neural system [3]. They are so many types of neural networks and every one of them represents a different aspect of recognition and prediction. The general neural network types of learning are supervised and unsupervised. Here is described Supervised [6].

Problem that we often face in the learning process is overfitting [2]. It appears in different situations, whenever the parameters become worse instead of getting better. There are different methods that help to avoid the overfitting – "Early Stopping" and "Regularization". The description of Early Stopping can be found below. [5].

When multilayer neural network is trained, usually the available data must be divided into three subsets. The first subset is Training set, which is used for computing the gradient and updating the network weighs and biases. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set typically begins to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned [4].

The test set error is not used during training, but it is used to compare different models. It is also useful to plot the test set error during the training process. If the error in the test set reaches a minimum at a significantly different iteration number than the validation set error, this might indicate a poor division of the dataset [4].

There are four functions provided for dividing data into training, validation and test sets. They are:

- dividerand (Fig. 1) – the data is randomly divided into the three subsets [4];

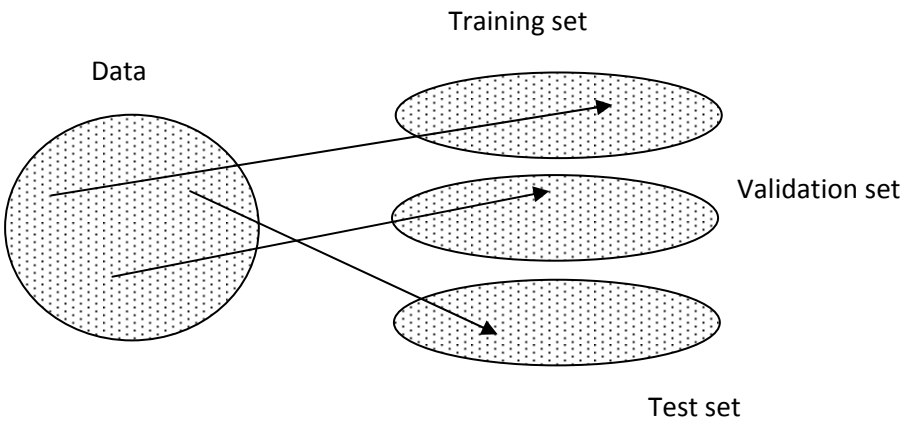

Figure 1.

- divideblock – the data is divided into three subsets using three contiguous blocks of the original data set [4];
- divideint – the data is divided by an interleaved method, as in dealing a deck of cards [4];
- divideind – the data is divided by index [4];

The next step after dividing the data is to train the network. The weights and biases are computed by using the training set only for one iteration. After that calculation mean square error and begins the process of verification. The training of the current network proceeds with validation set and calculation of the mean square error again and this operation gets repeated until the end with test set. If validation check reaches the maximum, the training process finishes and the mean square error and other parameters return in their values before the first overfit.

2. Generalized net model

Initially the following tokens enter the GN:

- In place L_1 – one token with initial characteristic “Input and target vectors”;
- In place L_8 – one token with initial characteristic “Methods of verification and the size of the data”;
- In place L_{12} – one token with initial characteristic “Structure of Neural Network”
- In place L_{13} – one token with initial characteristic “Maximum iterations, Maximum verification checks and minimum mean squared error”;
- In place L_{28} – one token with initial characteristic “Initial and maximum corrections”

The GN is presented on Fig. 2. by a set of transitions: $A = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8\}$. These transitions describe the following processes:

- Z_1 = “Preparing the data”
- Z_2 = “Appropriation maximum size of the data”
- Z_3 = “Appropriation methods of dividing and proportions’ size of the data”
- Z_4 = “Dividing the data”
- Z_5 = “Training of Neural network”
- Z_6 = “Verification of Neural network”
- Z_7 = “Reset validation check”
- Z_8 = “Correction of the verification”

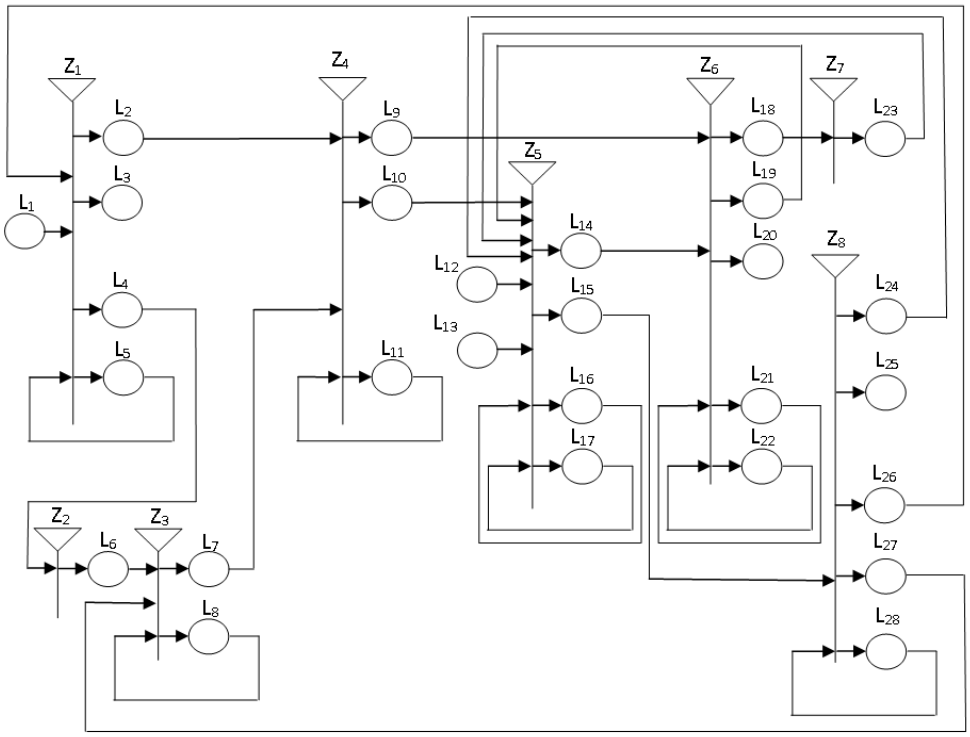

Figure 2. The GN model

The transitions are described below, as follows.

$$Z_1 = \langle \{L_1, L_5, L_{26}\}, \{ L_2, L_3, L_4, L_5\}, R_1, \vee(L_1, L_5, L_{26})\rangle,$$

where

$$R_1 = \begin{array}{c|cccc} & L_2 & L_3 & L_4 & L_5 \\ \hline L_1 & false & false & false & true \\ L_5 & W_{5,2} & W_{5,3} & W_{5,4} & true \\ L_{26} & false & false & false & true \end{array},$$

where

- $W_{5,2} = W_{5,4}$ = "Data were normalized"
- $W_{5,3}$ = "Data were not normalized and network stop"

The tokens enter place L_5 from place L_1 and do not obtain new characteristics. They stay in place L_5 for the life time of the GN.

The tokens enter places L_2 and L_4 from place L_5 and obtain a new characteristic "Normalized data".

$$Z_2 = \langle \{L_4\}, \{L_6\}, R_2, \vee(L_4) \rangle,$$

where

$$R_2 = \begin{array}{c|c} & L_6 \\ \hline L_4 & True \end{array},$$

The token enters in place L_6 from place L_4 and obtains a new characteristic "Maximum number of vectors".

$$Z_3 = \langle \{L_6, L_8, L_{27}\}, \{L_7, L_8\}, R_3, \vee(\wedge(L_6, L_8), L_{27}) \rangle,$$

where

$$R_3 = \begin{array}{c|cc} & L_7 & L_8 \\ \hline L_6 & False & True \\ L_8 & W_{8,7} & True \\ L_{27} & False & True \end{array},$$

where $W_{8.7}$ = "Method for dividing the data and the size of proportions were chosen"

The token from place L6 unites with token from L8 and obtains a new characteristic "Method of dividing were chosen; size of proportions; maximum number of vectors".

The token enters in place L_7 from place L_8 and does not obtain a new characteristic.

$$Z_4 = \langle \{L_2, L_7, L_{11}\}, \{L_9, L_{10}, L_{11}\}, R_4, \vee(\wedge(L_2, L_7), L_{11}) \rangle,$$

where

$$R_4 = \begin{array}{c|ccc} & L_9 & L_{10} & L_{11} \\ \hline L_2 & False & False & True \\ L_7 & False & False & True \\ L_{11} & W_{11,9} & W_{11,10} & True \end{array},$$

where $W_{11,9} = W_{11,10}$ = "Divided data"

The tokens transfer in place L_{11} from places L_2 and L_7 and unite in place L_{11} and stay there for the life time of the GN.

The token from place L_{11} splits between two equal token, where the first stays in place L_{11} and the second splits between L_9 and L_{10}, where they obtain new characteristics accordingly: “Verification data” and “Training data”.

$$Z_5 = \langle \{L_{10}, L_{12}, L_{13}, L_{16}, L_{17}, L_{19}, L_{23}, L_{24}\}, \{L_{14}, L_{15}, L_{16}, L_{17}\}, R_5, \vee(\wedge(L_{10}, L_{12}, L_{13}), L_{16}, L_{17}, L_{19}, L_{23}, L_{24})\rangle,$$

where

$R_5 =$

	L_{14}	L_{15}	L_{16}	L_{17}
L_{10}	*False*	*False*	*True*	*False*
L_{12}	*False*	*False*	*True*	*False*
L_{13}	*False*	*False*	*True*	*True*
L_{16}	*False*	*False*	*True*	$W_{16,17}$
L_{17}	$W_{17,14}$	$W_{17,15}$	*False*	*True*
L_{19}	*False*	*False*	*True*	*False*
L_{23}	*False*	*False*	*True*	*False*
L_{24}	*False*	*False*	*True*	*True*

where

- $W_{16,17}$ = “Mean square error, weights and bias were computed ”;
- $W_{17,15}$ = “Maximum number of iterations or minimum square error were reached”;
- $W_{17,14} = \neg W_{17,15}$.

The token from place L_{13} splits into two tokens accordingly in places L_{16} and L_{17}, where they obtain characteristics “Maximum number of iterations and minimum square error” and “Maximum number verification checks”.

The tokens enter place L_{16} from places L_{10}, L_{12} and L_{13} and obtain the characteristic “Computed weights, biases and mean square error”. Tokens that enter place L_{16} and stay there for the whole lifetime of the GN.

The token move to place L_{17} from places L_{13} and L_{16} and obtain characteristics “Current number of iterations”. The token in place L_{17} stays there for the whole lifetime of the GN.

The token from place L_{17} enters place L_{14} and obtains the characteristic “Current trained Neural Network”. The token from place L_{17} enters place L_{15} and obtains the characteristic “Trained Neural Network”.

$$Z_6 = \langle \{L_9, L_{14}, L_{21}, L_{22}\}, \{L_{18}, L_{19}, L_{20}, L_{21}, L_{22}\}, R_6, \vee(\wedge(L_9, L_{14}), L_{21}, L_{22})\rangle,$$

where

$$R_6 = \begin{array}{c|ccccc} & L_{18} & L_{19} & L_{20} & L_{21} & L_{22} \\ \hline L_9 & False & False & False & True & False \\ L_{14} & False & False & False & True & False \\ L_{21} & W_{21,18} & False & False & True & W_{21,22} \\ L_{22} & False & W_{22,19} & W_{22,20} & False & True \end{array},$$

where

- $W_{21,22}$ = “Neural Network is overfitting”;
- $W_{21,18} = \neg W_{21,22}$;
- $W_{22,19}$ = “Maximum validation check was not reach”;
- $W_{22,20} = \neg W_{22,19}$.

The tokens enter place L_{21} from places L_9 and L_{14} do not obtain a new characteristic.

The token enters place L_{18} from place L_{21} obtain a new characteristic “Verified Neural Network”.

The token enters place L_{22} from place L_{21} obtain a new characteristic “Number of verification checks”. The tokens in place L_{22} stay there for all life time of the GN.

The token enters place L_{19} from place L_{22} obtain a new characteristic “Overfitted Neural Network”.

The token enters place L_{20} from place L_{22} obtain a new characteristic “Maximum validation checks”.

$$Z_7 = \langle \{L_{18}\}, \{L_{23}\}, R_7, \vee(L_{18}) \rangle,$$

where

$$R_7 = \begin{array}{c|c} & L_{23} \\ \hline L_{18} & True \end{array}.$$

The token enters in place L_{23} from place L_{18} and obtains a new characteristic “Current validation check becomes zero”.

$$Z_8 = \langle \{L_{15}, L_{28}\}, \{ L_{24}, L_{25}, L_{26}, L_{27}, L_{28}\}, R_8, \vee(L_{15}, L_{28}) \rangle,$$

where

$$R_8 = \begin{array}{c|ccccc} & L_{24} & L_{25} & L_{26} & L_{27} & L_{28} \\ \hline L_{15} & False & False & False & False & True \\ L_{28} & W_{28,24} & W_{28,25} & W_{28,26} & W_{28,27} & True \end{array},$$

where

- $W_{28,25}$ = “Cannot improve characteristics of the Neural Network”
- $W_{28,24} = W_{28,26} = W_{28,27} = \neg W_{28,25}$

The token enters in place L_{28} from place L_{15} and obtains a new characteristic “Number of data corrections”. The tokens in place L_{28} stay there for the life time of the GN.

The tokens enter places L_{24}, L_{26} and L_{27} from place L_{28} and obtain a new characteristic "Corrections of data and characteristics".

3. Conclusion

One of the verification methods was described by the generalized nets model. Its name is "Early Stopping". The main function of this method is to interrupt the learning of the neural network when it overfits. This is the second paper devoted to Neural Network over the process of verification. Other methods and algorithms will be an object of future developments.

References

[1] Atanassov, K., *Generalized Nets*, World Scientific, Singapore, 1991.

[2] Geman, S., E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma, *Neural Computation*, Vol. 4, 1992, 1–58.

[3] McCulloch, W., W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, Vol. 5, 1943, 115–133.

[4] Beale, M. H., M. T. Hagan, H. B. Demuth, *Neural Network Toolbox User's Guide R2012a*, 1992–2012.

[5] Morgan, N., H. Bourlard, Generalization and parameter estimation in feedforward nets: some experiments. In: *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann Publishers Inc. San Francisco, 1990, 630–637.

[6] Widrow, B., M. Hoff, *Adaptive Switching Circuits*, IRE WESCON Convention Record, 1960, 96–104.