



Digital Receipt

This receipt acknowledges that **Turnitin** received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Sumarno Linggo
Assignment title: Periksa similarity
Submission title: Generating A Combination of Backpropagation Neural Netw...
File name: pagation_Neural_Networks_by_Means_of_Random_Seeds_Ev...
File size: 343.96K
Page count: 6
Word count: 3,532
Character count: 19,010
Submission date: 07-Feb-2018 02:43PM (UTC+0700)
Submission ID: 912404476



Generating A Combination of Blackpropagation Neural Networks by Means of Random Seeds Evaluation

by Linggo Sumarno

Submission date: 07-Feb-2018 02:43PM (UTC+0700)

Submission ID: 912404476

File name: pagation_Neural_Networks_by_Means_of_Random_Seeds_Evaluation.pdf (343.96K)

Word count: 3532

Character count: 19010

Generating A Combination of Backpropagation Neural Networks by Means of Random Seeds Evaluation

2
Linggo Sumarno

Faculty of Science and Engineering
Sanata Dharma University, Yogyakarta 55282, Indonesia.
Email: lingsum@dosen.usd.ac.id

ABSTRACT

5
A classifier combination is becoming a new trend in pattern recognition and machine learning. The development of new techniques, in order to generate a combination of several classifiers is an important aspect in a classifier combination. Backpropagation neural network is a kind of classifier. Usually, a combination of backpropagation neural networks is generated by using different architectures i.e. different number of layers and also different number of neurons in a layer. Based on the experiment, it was shown that a combination of backpropagation neural networks could also be generated by using the same architecture but different weight sets. Those neural networks were generated by means of random seeds evaluation when they were trained.

Keywords

Combination, backpropagation neural network, random seed

1. INTRODUCTION

3
In early studies of pattern recognition, only one classifier was used to solve a classification problem. In the early 1990's, an idea emerged that in a pattern recognition not only one classifier but also several classifiers could be used. In accordance with it, the idea to use classifier combination methods has been expanding. The research domain of classifiers combination methods examine how several classifiers can be applied together to obtain better classification systems. It was shown that classifier combination methods might improve the recognition performance in difficult pattern recognition problems [9], [10]. Classifier combination methods may also be used to increase the speed of the systems [2], [12] or to reduce the time taken for the design of classification system [4].

3
There are two main issues in classifier combination methods. The first one is how the individual classifiers are generated, and the second one is how those classifiers are combined. This paper will discuss the first issue, i.e. how a classifier combination can be generated based on a simple concept. In this paper, a combination of backpropagation neural networks (where several backpropagation ones have the same architecture but different weight sets) has been studied experimentally.

2. THEORY

2.1 Classifier Combination

Classifier combination methods have proved to be an effective tool to improve the performance of pattern recognition applications. In terms of classifier combination members, theoretical research by Hansen and Salomon [5] and also Krogh and Vedelsby [11], as well as empirical research by Hashem [6] and Optiz [16] have demonstrated that the best combination is combination of several different classifiers. There were no advantages in combining several identical classifiers.

In order to generate several different classifiers above, it can be carried out by only based on a base classifier. This generation can be carried out by changing the training sets [1], changing the input features [8], [16], or changing the parameters and architecture of a base classifier [17].

2.2 Backpropagation Neural Network

A neural network defined as a computational structure that consists of parallel interconnections neural processors, which have adaptation ability. Backpropagation neural network is a neural network that commonly used. Figure 1 shows an example of a backpropagation neural network that used in this research. It consists of C_0 input unit, C_1 and C_2 neurons in the hidden layer 1 and 2 respectively, and also C_3 neurons in the output layer.

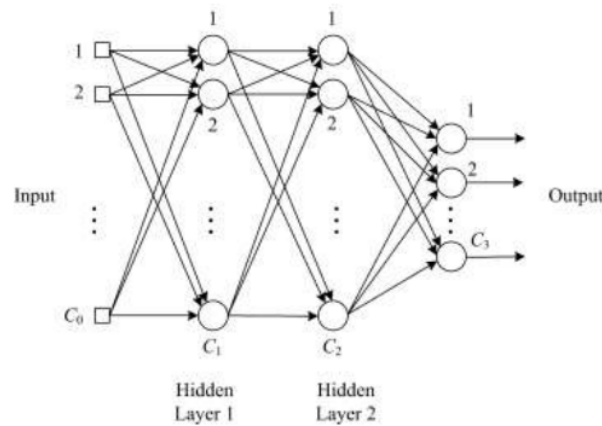


Figure 1: An example of a backpropagation neural network with two hidden layers.

2.2.1 Training

In order that neural network in Figure 1 can be used in recognizing the applied input, it needs to be trained first. Resilient propagation [18] is one of many training algorithms to train the backpropagation neural network.

2.2.2 Initial Weights

One thing that carried out during the early step neural network training is assigning the initial weights of the neurons. The choice of the weights will influence the convergence rate or even the convergence failure of the neural network training.

1. If the initial weights are set at the same values, the resulted error values will be constant over all training period. This situation will cause the neural network training trapped in the saturation that resist weights changing. Therefore, it can be judged that a convergence failure has been happened.
2. If the initial weights are set at the different values (however they are inappropriate), it will cause a phenomenon called premature saturation [13]. This phenomenon refers to a situation where the resulted error values almost constant over a training period. This phenomenon cannot be judged as a local minimum, since error value will be reduced after that almost constant period over. This premature saturation phenomenon will slow down the convergence rate.

In order to avoid the two things above, in general practice, researchers used initial weights from random numbers that uniformly distributed and also in the small range [7].

2.2.3 Random Numbers and Random Seeds

One of computer property is deterministic property. Therefore, it cannot generate the real random numbers. Computer uses a pseudorandom generator, in order to mimic the real random number generator. By using this kind of generator, it can be generated a series of exact pseudorandom numbers, as long as the generator is initialized using the same initial number. This initial number called random seed.

When a process that make use a series of pseudorandom numbers is executed, it is possible to get an identical track record of the process. The neural network also makes use a series of pseudorandom numbers in the training process. Therefore, it is possible to get an identical track record of the training process, even though the training process is repeated again. On the other hand, by using a different series of pseudorandom numbers in the training process, it is possible to get a different track record of the training process. In the neural network training, a different track record means a different neural network, since it will has a different performance.

3 RESEARCH METHODOLOGY

2 Materials and Equipments

Materials in this research are isolated handwritten words and characters in binary format. These materials came from data acquisition sheets scanned at resolution of 300 dpi. The data were taken from 100 writers, from various levels of age (10-70

years) and sex. From 100 writers, each of them wrote 78 characters, which divided into three groups where each group consists of 'a'-'z' characters. Therefore, there were 7,800 isolated character images.

Equipments in this research was a set of computer equipped by processor Intel Core2Duo E7500 (2,93GHz) and 2GB RAM, that consists of MATLAB 7.0.4.365 (R14) software.

3.1 System Development

By using materials and equipments above, a system of handwritten character recognition has been developed (see Figure 2). In that system, the input is an isolated character image in binary format, whereas the output is a character in the text format.

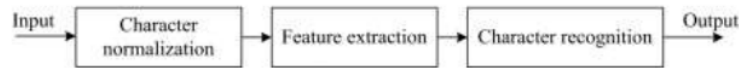


Figure 2: A character recognition system.

3.1.1 Character Normalization

Character normalization in Figure 2 carried out in order to correct problems of slant, size, shift, and stroke-width. In this research character normalization from Sumarno [21] was used. Figure 3 shows some steps in character normalization.

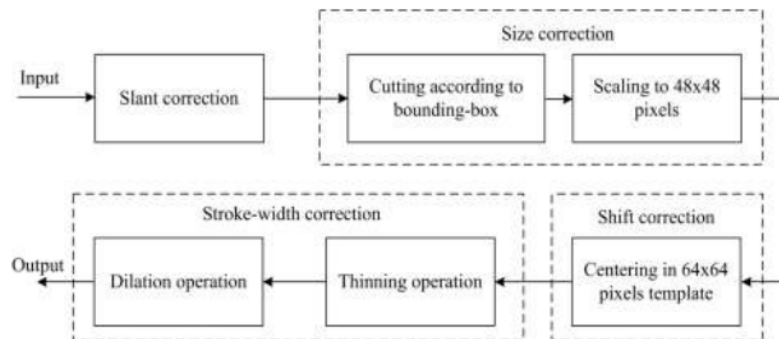


Figure 3: A character normalization steps.

In Figure 3 the input is an isolated handwritten character in binary format, whereas the output is normalized handwritten character in binary format which has 64x64 pixels in size. Slant and size correction made use of linear transform of shearing and scaling respectively. Stroke-width correction made use of morphological operations i.e. thinning and dilation. Shift correction made use of a simple method by placing the character image in the center of the used template. Sumarno [21] suggested the following parameters.

1. Slant correction was carried out by using evaluation of vertical projection histogram of handwritten character that had been undergone shearing operation by using shearing coefficients $\{-0.4, -0.35, \dots, 0.4\}$ (In this case, it was assumed that the slant of handwritten character was in the range of shearing coefficient -0.4 to 0.4). The straightness of the character corresponds with a shearing coefficient that gives highest variance.
2. Character scaling was set to 48x48 pixels.
3. The template size was set to 64x64 pixels.
4. Thinning operation used thinning algorithm from Zhang-Suen [23].
5. Dilation operation used square structure-element 3x3.

10

3.1.2 Feature Extraction

Feature extraction is a process to extract features that exist in each of the character image. In this research feature extraction from Sumarno [21] was used. Figure 4 shows some steps in feature extraction.

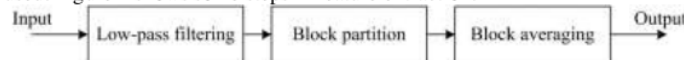


Figure 4: Feature extraction steps.

In Figure 4 above, the input is normalized character in binary format that has 64x64 pixels in size. The output is a set of values that represents the input image that has 64 elements. The aim of low-pass filtering is to blur the input image. The aim of block partition is to partition the image into blocks of image, for block averaging needs. The aim of block averaging is to get a set of values that represents the input image. Sumarno [21] suggested the following parameters for the feature extraction steps.

1. Low-pass filtering used 2D Gaussian filter 14x14 with standard deviation 10.
2. Block partition used 8x8 pixels block size.

3.1.3 Character Recognition

Character recognition is a process that needed to recognize a character, based on the trained character features. In order to recognize it, a recognition method based on a backpropagation neural network is used. This kind of neural network also known as Multi Layer Perceptron (MLP), which introduced by Rosenblatt [19] and developed by Minsky and Papert [14], [15]. Backpropagation neural networks that used in this research described in detail as follow.

1. Input layer has 64 neurons that correspond with the number of feature extraction elements.
2. Output layer has 26 neurons that correspond with the number of alphabet characters 'a' to 'z'. Transfer function in this layer is unipolar sigmoid, that correspond with the network output i.e. in the range of 0 to 1.
3. Neural network has 2 hidden layers i.e. hidden layer 1 and 2 which have 64 and 312 neurons respectively. The number of neurons in each hidden layer was found from an evaluation procedure, where by using 64 and 312 neurons in hidden layer 1 and 2 respectively, it gave the highest recognition rate. Transfer functions in each hidden layer is bipolar sigmoid, that correspond with internal data processing in neural network which in the range -1 to 1.

Notes:

1. In case of pattern recognition that based on multiresolution, Suhardi [20] found that a backpropagation neural network with two hidden layers, could give better recognition rate compare with one hidden layer.
2. Sigmoid function is a function that commonly used in a backpropagation neural network [3].
3. Training of a backpropagation neural network can be more effective by using bipolar data processing in the range of -1 to 1 [20].

Training algorithm

The neural network trained by using resilient backpropagation algorithm [18]. This algorithm is the fastest algorithm for pattern recognition [13]. Stopping criterion in training made use of validation, in order to avoid under-training or over-training.

Pseudorandom numbers

Since the computer cannot generate the real random numbers, therefore the pseudorandom numbers were used in the neural network as initial weights. That pseudorandom numbers have the following properties.

1. Distribution : uniform (see subsection 2.2.2)
2. Range of value : -1 to 1 (since bipolar sigmoid function has limit numbers -1 and 1)
3. Repeatability : 2^{1492} (built-in in the MATLAB software)

In order to remove correlation between layers, initial weights in each layer should be different. Therefore, random seed value that used in generating pseudorandom numbers in each layer should be different. In this research, random seed values that used in hidden layers 1, 2 and output layer were i (i are integer number), $i+1$ and $i+2$ respectively.

Patterns in training and testing

Patterns that used in training and testing the neural network are images of isolated handwritten character, which come from 100 persons that further processed into three pattern sets as follows.

1. **Training set**
This set used in training (in updating the neuron's weights). This set consists of 13,000 patterns as follows.
 - a. There are 2,600 corrected patterns from group 1.
 - b. There are 5,200 corrected patterns from group 2. They come from corrected patterns from group 2 that rotated -5^0 and 5^0 .
 - c. There are 5,200 corrected patterns from group 3. They come from corrected patterns from group 3 that rotated -10^0 and 10^0 .

Notes:

- a. Every group consists of 2,600 patterns.
 - b. Corrected patterns are original patterns that have undergone slant, size, shift, and stroke width corrections.
 - c. It was assumed that the rotation in input patters is in the range of -10^0 to 10^0 .
2. **Validation set**
This set also used in training (in stopping the training process). They consist of 2,600 corrected patterns from group 2.
 3. **Test set**
This set used in testing the trained neural network. They consist of 2,600 corrected patterns from group 3.

4. TESTING AND DISCUSSIONS

In generating a combination of neural networks, firstly several neural networks were trained by using the different random seeds but the same training sets. In this research, 10 neural networks were trained by using random seeds 1 to 10. Table 1 shows the training and testing results.

Table 1: Random seeds evaluation in generating 10 neural networks.

Random seed values	Number of epochs when training stopped	Value of MSE when training stopped	Character recognition rate (%)
1	474	0.00155	87.39
2	360	0.00203	87.15
3	376	0.00216	85.00
4	410	0.00196	86.69
5	383	0.00235	85.00
6	436	0.00196	86.38
7	345	0.00241	85.62
8	533	0.00144	86.69
9	402	0.00188	86.31
10	351	0.00211	86.46

Notes:

1. MSE : Mean Square Error
2. All neural networks are backpropagation neural networks that have the same architecture i.e. 64-64-312-26.
3. The above random seed values are used in the first hidden layer. In the second and output layers, the random seed values are $i+1$ and $i+2$ respectively, where i is a random seed value in the first hidden layer.

Table 1 above shows that random seed values that used in setting initial weights of 10 neural networks, have effects in character recognition rates. This case was happened due to by using different weight sets, the training of neural networks started from different starting points. By starting from different starting points, the resilient backpropagation training gave different epochs (and different MSE) when training stopped. Finally, it will be obtained several neural networks that have the same architecture but different weight sets. As shown in Table 1, several neural networks that have the same architecture but different weight sets have different performance in terms of character recognition rate. However, Table 1 also shows that there are some random seed values, which give the same character recognition rates, i.e. random seed values 3 and 5, and also 4 and 8 give recognition rates 85% and 86.69% respectively. Although they have the same recognition rate, Table 1 shows that random seed values 3, 4, 5, and 8 have different number of epochs and MSE (it means they are different neural networks). Therefore, it can be said that the different neural networks may give the same recognition rates.

Once several neural networks that will be used in a combination of neural networks have been generated, a sorting procedure can be carried out. Table 2 shows sorting result of Table 1 that based on recognition rates. The sorting procedure was carried out in order to easily choose which neural networks would be used. For example, if the combination will use five neural networks, then the best five can be chosen easily.

As discussed in the subsection 2.1, in order to generate several different classifiers, which will be used in a classifier combination, it can be carried out by only based on a base classifier i.e. by changing the training sets [1], changing the input features [8], [16], or changing the parameters and architectures of a base classifier [17]. In this research, generating several different classifiers was carried out by based on simplification of Partridge and Yates concept [17], i.e. by only changing the parameters of a base classifier.

Table 2: Neural network sorting based on character recognition rate.

Neural network number	Character recognition rate (%)	Random seed values
1	87.39	1
2	87.15	2
3	86.69	4
4	86.69	8
5	86.46	10
6	86.38	6
7	86.31	9
8	85.62	7
9	85.00	3
10	85.00	5

5. CONCLUSION

Based on the above description, there are two conclusions as follow.

1. A new study about a combination of backpropagation neural networks has been carried out. This combination consists of backpropagation neural networks that have the same architecture but different weight sets. Those neural networks were generated by means of random seeds evaluation when they were trained using the same training set.
2. Generation of combination members by a simpler concept, i.e. by means of random seed evaluation has been studied experimentally. That generation concept is simpler than the other reported generation concepts, like changing the training sets [1], the input features [8], [16], or the parameters and the architectures of a base classifier [17].

REFERENCES

- [1] L. Breiman, "Bagging Predictors", *Machine Learning*, vol. 24, no. 2, pp 123 – 140, 1996.
- [2] Y. Chim, A. Kasim, and Y. Ibrahim, "Dual Classifier System for Handprinted Alphanumeric Character Recognition", *Pattern Analysis and Applications*, vol. 1, pp. 155-162, 1998
- [3] L. Fausett, *Fundamentals of Neural Networks*, Prentice Hall International Inc., New Jersey, 1994.
- [4] S. Gunter, *Multiple Classifier Systems in Offline Cursive Handwriting Recognition*, PhD Thesis, Institute for Computer Science and Applied Mathematics, University of Bern, German, 2004.
- [5] L. Hansen, and O. Salomon, "Neural Network Ensembles". *IEEE Transactions on Pattern Recognition and Machine Analysis*, vol 12, no. 10, pp. 993-1001, 1990.
- [6] S. Hashem, "Optimal Linear Combinations of Neural Networks", *Neural Networks*, vol. 10, no. 4, pp. 599-614, 1997.
- [7] S. Haykin, *Neural Networks : A Comprehensive Foundation*, Prentice-Hall International Inc., New Jersey, 1994.
- [8] T.K. Ho, "The Random Supspace Method for Constructing Decision Forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp.832 – 844, 1998.
- [9] J. Kittler and F. Rolli, editors, *Proceeding of the First International Workshop on Multiple Classifier Systems*, Cagliari, Italy, 2000.
- [10] J. Kittler and F. Rolli, editors, *Proceeding of the Second International Workshop on Multiple Classifier Systems*, Cambridge, UK, 2001.
- [11] A. Krogh, and J. Vedelsby, "Neural Networks Ensembles, Cross Validations, and Active Learning", *Advances in Neural Information Processing Systems*, G. Tesauro et.al. (editor), MIT Press, vol. 7, pp. 231-238, 1995.
- [12] L.Lam and C. Y. Suen, Structural Classification and Relaxation Matching of Totally Unconstrained Handwritten Zip-code Numbers, *Pattern Recognition*, vol. 21, no. 1, pp. 19-31, 1998.
- [13] Y. Lee, S. Oh, and M.Kim, "The Effect of Initial Weights on Premature Saturation in Backpropagation Learning". *International Joint Conference on Neural Networks*, Seattle, Washington, USA, vol. 1, pp.765-770, 1991.
- [14] M. L. Minsky and A. S. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.
- [15] M. L. Minsky, and A. S. Papert, *Perceptrons, Expanded Edition*, MIT Press, 1988.
- [16] D. W. Optiz, "Feature Selection for Ensembles", *Proceeding of 16th International Conference on Artificial Intelligence*, pp. 379-384, 1999.
- [17] D. Partridge, and Y. B. Yates, "Engineering Multiversion Neural-Net Systems", *Neural Computation*, vol. 8, no. 4, pp. 869-893, 1996.
- [18] M. Riedmiller, and H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm", *Proceedings of the IEEE International Conference on Neural Networks*, pp. 586-591, 1993.
- [19] F. Rosenblatt, *Principles of Neurodynamics*, Spartan, New York, 1962.
- [20] I. Suhardi, *Evaluation of Artificial Neural Network for Handwritten Character Recognition Handprinted Style* (in Indonesian), Master Thesis, Electrical Engineering, Gadjah Mada University, Yogyakarta, 2003.
- [21] L. Sumarno, "On The Performance of Blurring and Block Averaging Feature Extraction Based On 2D Gaussian Filter". *Proceeding of 5th International Conference on Information and Communication Technology and Systems (ICTS)*: Surabaya, Indonesia, pp. 261-266, 2009.
- [22] The Mathworks Inc., *Neural Network Toolbox: For Use with MATLAB*, Version 5, The Mathworks Inc., Massachussets, 2005.
- [23] T.Y. Zhang, T.Y. and C. Y. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns", *Comm. ACM*, vol. 27, no. 3, pp. 236-239, 1984.

Linggo Sumarno

Received S1 (1991), S2 (1999), and S3 (2010) degrees from Electrical Engineering Department of Gadjah Mada University, Yogyakarta, Indonesia. Currently he is a lecturer in Electrical Engineering Department of Sanata Dharma University, Yogyakarta, Indonesia. His research interests are signal processing, image processing, and pattern recognition.

Generating A Combination of Blackpropagation Neural Networks by Means of Random Seeds Evaluation

ORIGINALITY REPORT

17%

SIMILARITY INDEX

10%

INTERNET SOURCES

12%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Universitas Jember

Student Paper

4%

2

Sumarno, Linggo. "Handwritten word segmentation using Kaiser window", 2013 International Conference on QiR, 2013.

Publication

3%

3

www.iam.unibe.ch

Internet Source

3%

4

Gunter, S.. "Off-line cursive handwriting recognition using multiple classifier systems- on the influence of vocabulary, ensemble, and training set size", Optics and Lasers in Engineering, 200503/05

Publication

1%

5

Studies in Computational Intelligence, 2008.

Publication

1%

6

Parthasarathy Subhasini, Bernadetta Kwintiana Ane, Dieter Roller, Marimuthu Krishnaveni. "chapter 10 Intelligent Classifiers

1%

Fusion for Enhancing Recognition of Genes and Protein Pattern of Hereditary Diseases", IGI Global, 2012

Publication

7	link.springer.com Internet Source	1 %
8	Wengui Rao, Yan Dong, Shaoping Chen, Fang Lu, Shu Wang. "An Efficient Rate Compatible Modulation With Variable Weight Sets", IEEE Access, 2018 Publication	<1 %
9	www.hindawi.com Internet Source	<1 %
10	www.mdpi.com Internet Source	<1 %
11	Artificial Neural Nets and Genetic Algorithms, 1995. Publication	<1 %
12	Lecture Notes in Computer Science, 2004. Publication	<1 %
13	hdl.handle.net Internet Source	<1 %
14	Ruta, D.. "Classifier selection for majority voting", Information Fusion, 200503 Publication	<1 %
15	cs.nyu.edu	

Internet Source

<1 %

16

hrcak.srce.hr

Internet Source

<1 %

17

Aircraft Engineering and Aerospace
Technology, Volume 77, Issue 1 (2006-09-19)

Publication

<1 %

Exclude quotes On

Exclude matches < 5 words

Exclude bibliography On