Improving BP Algorithm in Music Classification

[1]Wei Gao, [2]Yadong Yang and [3]Yi Li

Abstract

By the further study of the variable learning rate methods on BP algorithm, this paper came up with the Segmented Learning Algorithm（SLA）for music classification where the music comes from CDs. Classification results are shown on a three-dimensional image. The accuracy of classification results is 90.08% by using the segmented learning algorithm.

Keywords: BP algorithm, Segmented Learning Algorithm, three-dimensional image, Music Classification

1. Introduction

Since 1985 Rumelhart and his colleagues developed a learning algorithm for the exploration, the study of neural networks has never stopped. The research and applications of an artificial neural network have been penetrated into various fields, especially in pattern recognition, classification, computer vision, adaptive filtering and signal processing, continuous music recognition. Variants of the back-propagation algorithm as well as unsupervised methods by Geoff Hinton and colleagues at the University of Toronto can be used to train deeply, and highly nonlinear neural architectures similar to the 1980 Neocognitron by Kunihiko Fukushima, and the "standard architecture of vision", were inspired by the simple and complex cells identified by David H. Hubel and Torsten Wiesel in the primary visual cortex.

*Corresponding Author: Wei Gao
(E-mail: gwe_mail@syuct.edu.cn)
[1]Shenyang University of Chemical Technology

Music classification is a problem of pattern recognition, including two aspects: feature extraction and classification. Traditional music classification methods, rule-based audio classification, and HMM [1] have their own deficiencies. Artificial neural networks provide a new way to solve the problem of music classification [2].

Rumelhart and McCelland have carried out a detailed analysis of BP algorithm and realization of Minsky's idea about multi-layered network[3]. In recent years, to solve the slow convergence of BP neural network and instability of network learning people proposed many improved algorithms, including additional momentum method and variable learning rate method [5]. The variable learning rate methods can obtain optimum values. However, the stability is still not significant. Here proposed the Segmented Learning Algorithm, an improvement of BP algorithm. Traditional BP algorithm uses a two-dimensional graph to display results, so the results is not intuitive. This paper shows the 3-D representation of the classification results.

2. Techniques Used in Music Classification

2.1 BP Algorithm in Music Classification

The operations of the BP neural networks can be divided into two steps: feedforward and Back Propagation. In the feedforward step, an input pattern is applied to the input layer, so its effect propagates layer by layer through the network until an output is produced. The network's actual output value is then compared with the expected output, and an error signal is computed for each of the output nodes. The output error signals are transmitted backwards from the output layer to each node in the hidden layer immediately contributed to the output layer, yet this process is then repeated. Once the error signal for

each node has been determined, the errors are then used by the nodes to update the values for each connection weights until the network converges to a state that allows all the training patterns to be encoded. The Back propagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent [4]. The weights that minimize the error function is then considered to be a solution to the learning problem. Music classification system is an important aspect in the field of pattern recognition. Music classification is generally divided into two steps: the learning phase and the recognition phase. The mission of learning phase is to establish the basic unit of acoustic models to identify the voice models. The input stage is to identify the target music feature parameters and compares the model to obtain a recognition result.

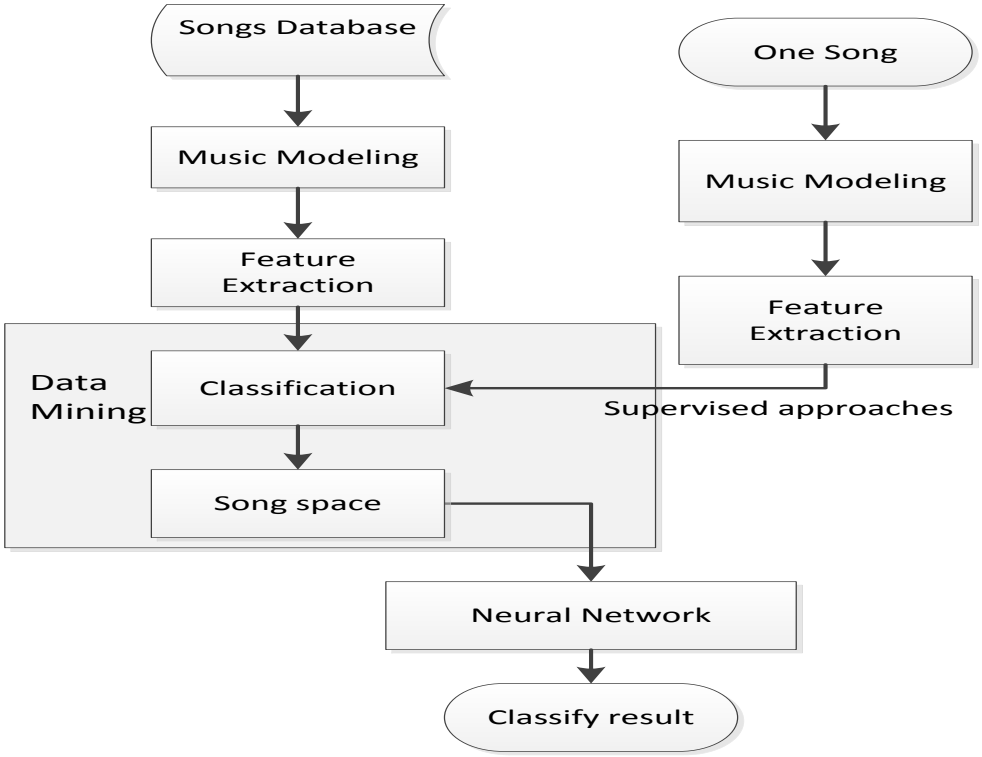

Figure 1: Supervised approaches of BP algorithm in Music classification

2.2 Segmented Learning Algorithm (SLA)

Variable learning rate method refers to the learning rate of early learning BP neural network which is large with rapid network convergence. Along with the learning process learning rates continue to decrease, the network becomes stabilized. Experimental results show that one-time reduction in the learning rate is not enough to stabilize the results, so learning rate, segmented learning and repeating need to be improved by the following steps:

Step 1: Set numbers of segments with given initial learning rate, and start the first stage of learning.

Step 2: When the learning rate drops to a relatively low value, reset a learning rate and give a value less than the initial learning rate, starting the next stage of learning. Repeat this step until the final period of study.

Step 3: Learn the final period until the learning rate has dropped to zero, and end learning process.

SLA algorithm flowchart is shown below:

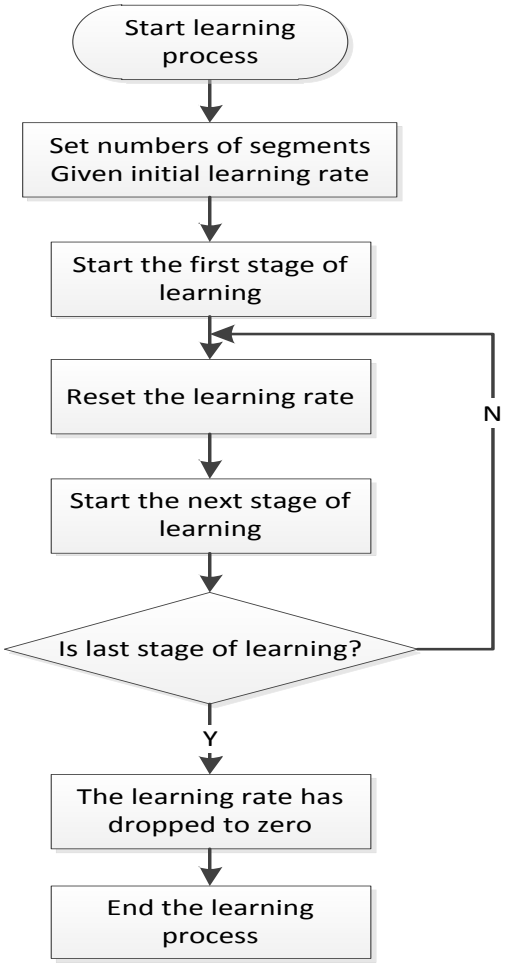

Figure 2: Segmented Learning Algorithm flowchart

Because many factors are artificially given, the optimal value cannot be accurately selected. For example, by several experiments, in this paper initial learning rate is 0.3, the number of segments are 3, the process of learning rate changed as shown below:

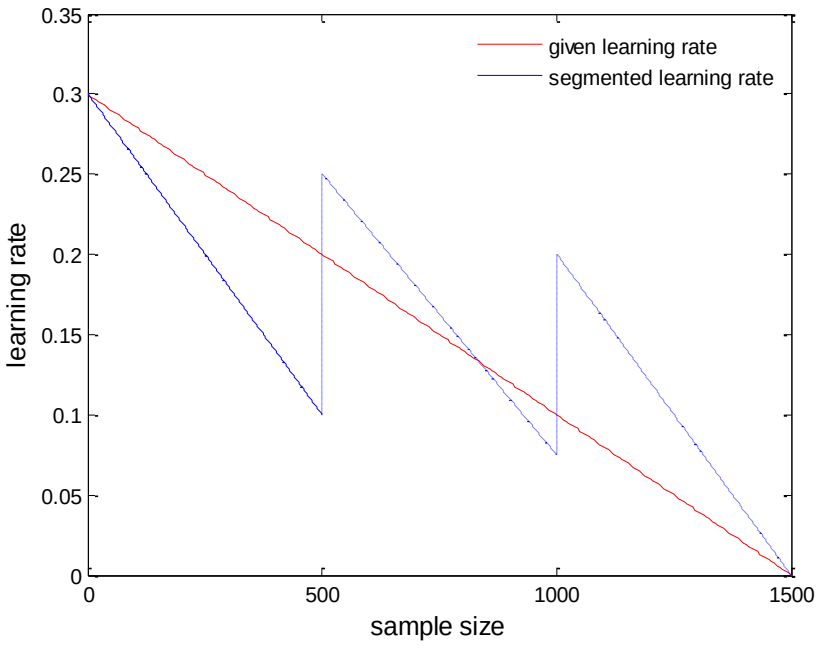

Figure 3: the learning rate of the Segmented Learning Algorithm

3. Feature Selection

Feature selection is important for audio classification. The selected features should reflect the significant characteristics of different kinds of audio signals. In order to distinguish different classes of audio, we consider the features related to temporal and spectral domains [7].

1). Song's duration in seconds

2). Tempo in beats per minute (BPM)

3). Root mean square (RMS) amplitude set in dB. The RMS value is the effective value of the total waveform. It is really the area under the curve. In audio it is the continuity or music power that the amplifier can deliver.

4). Sampling frequency in kHz and Sampling rate set in b. There are two major common sample rate bases, 44.1 kHz and 48 kHz. As most studio equipment uses recording, and digital signal processing equipment works at a sample rate based on 48 kHz; the final result has to be re-sampled to the standard CD-format sample rate finally transferred onto a CD for distribution.

5). Dynamic range (dr) set in dB. Dynamic range is the ratio between the largest and smallest possible values of a changeable quantity, such as in signals like sound and light.

6). Genre name: ZITHER, ROCK, FOLK, POP

Data used for training and testing the system was taken from 20 compact discs composed of 5 classified as folk, 5 classified as pop, 5 classified as zither, and 5 classified as rock. The tracks on each of these CDs were extracted and converted to WAV format, and then divided into segments of length 2^{18} bits. To avoid periods within the music without characteristic of the whole song, the segments were all taken from the middle of each track. This procedure was able to produce 2,683 segments of music. The segments of music were then further divided into two subs - segments by extracting the first 2^{16} bits and the third 2^{16} bits [8].

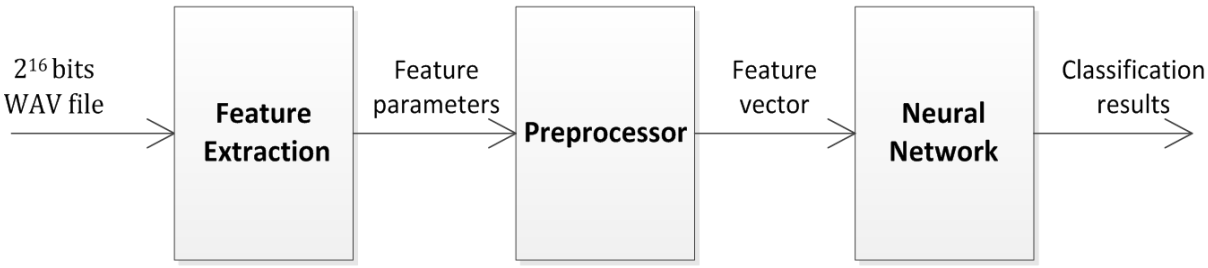

Figure 4: Digital Music Classification System

4. Results and Analysis

4.1 Classification Results

To test the performance of the music classification system, the system was first configured to classify music by genre. The four genres used were Zither, Rock, folk, and Pop. The first step in performing this test was to generate the data set. As discussed above, the data set was taken from 20 CDs with five per genre, and consisted of 2000 samples. Before training, data preprocessing was performed on the training data as discussed above. After preprocessing, the training data was divided further into two groups, one for training and the other for testing. From these 2000 sample data, 1500 were used for training, and the other 500 were reserved for testing [6]. A test data set was needed to ensure that the neural network did not over fit the data.

Here are the different outputs based on different categories of content, so we can see the number of classification errors and errors assigned to which category. From the analysis a higher success rate was obtained in classifying music by BP algorithm. The first category ZIRHER was classified 90% correctly while the second category ROCK got 97.74% success in classification. The success rates for the third category FOLK and fourth category POP were 81.36% and 100%, respectively. To get a more accurate result, you can zoom in Figure 3.

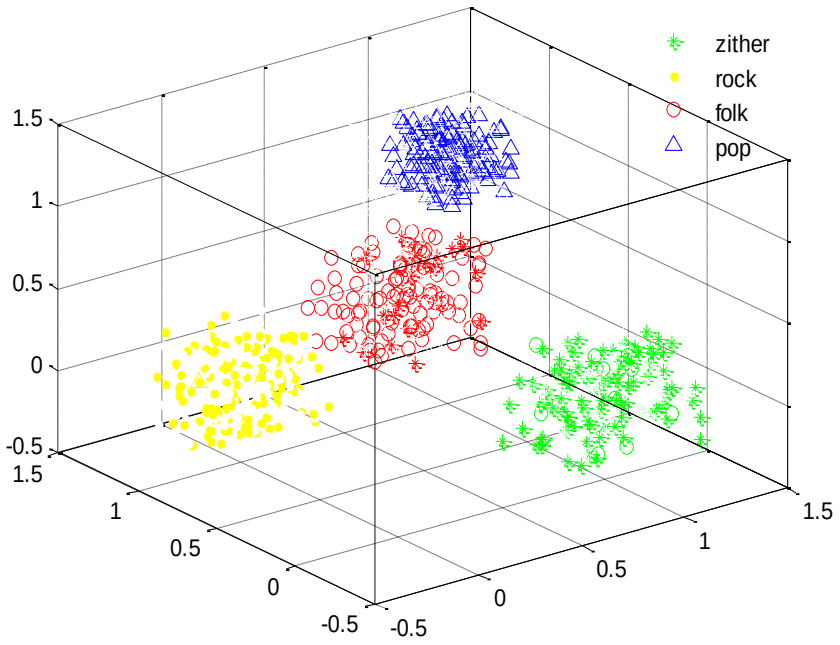

Figure 5: Classification results in SLA

Now, put the original music classification outputs and the misclassification outputs after training on the same image, so it is easy to see four kinds of misclassification. Four different shapes of the same color represent four types of original samples, and the shapes of the other four marked with different colors denote the misclassification outputs after training, as shown in Figure 4.

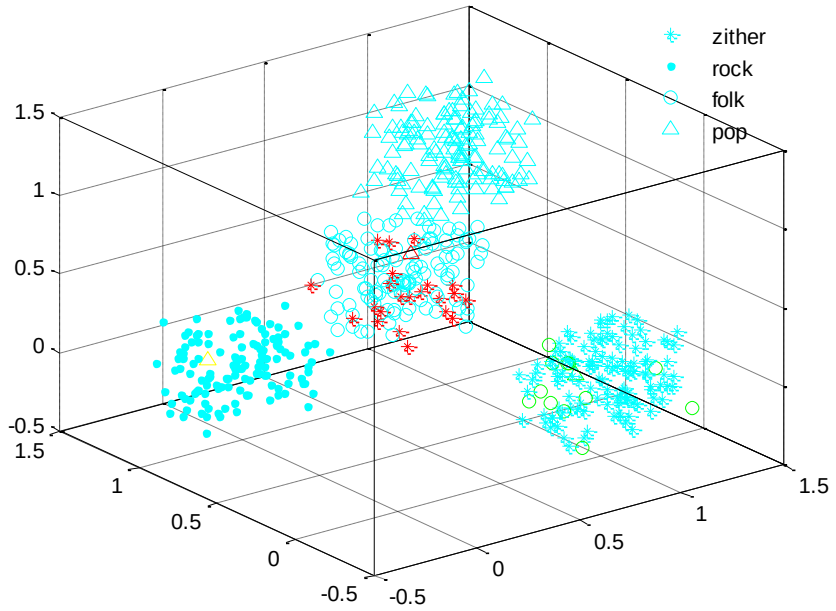

Figure 6: Show the misclassification

Respectively given learning rate, variable learning rate and segmented learning rate were conducted with several tests, and overall average results of each test are shown below:

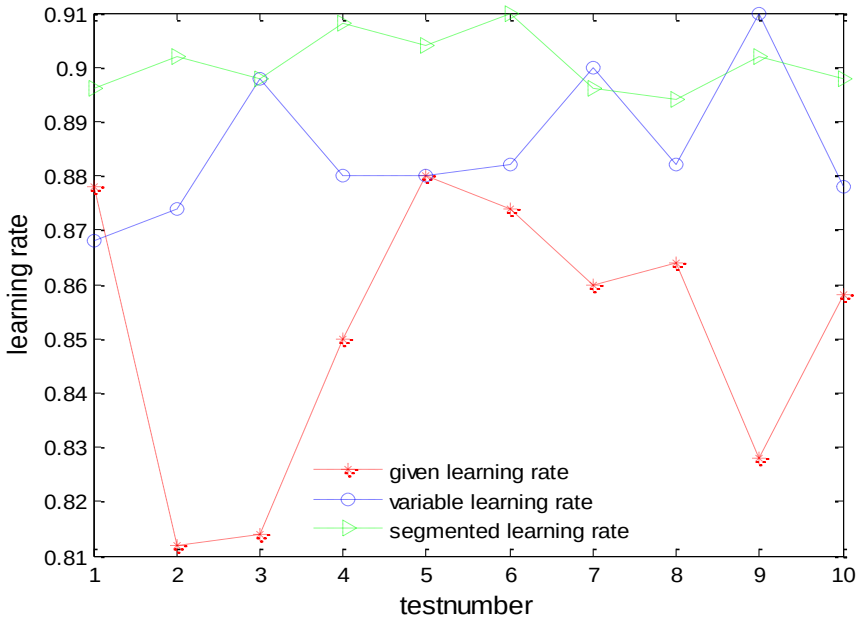

Figure 7: Comparison of the results obtained by the three methods

Figure 5 shows that the classification results obtained by the given learning rate method is not very good, while the correct rate for each classification is quite different. Classification results obtained by the variable learning rate method is significantly better than that by using the given learning rate method. And using Segmented Learning Algorithm better classification result and more stable network are obtained.

4.2 Results Analysis

The reason for this classification difference is mainly the instrument. For pop and rock the instruments used are quite same in all over the world. So their features have been judged more correctly than others. For the zither and folk songs less success was obtained compared with the other. The main reason is that the instrument used to play these two categories are different in many countries. Therefore, there are a few problems in extracting these two categories. The results obtained by Segmented Learning Algorithm method is better than the other two learning rate methods. The reason is that using Segmented Learning Algorithm method, the classification process is not easy to fall into a local optimum, yet this repeat memory's method is closer to the way of human memory.

5. Conclusion

Traditional BP algorithm is an outstanding neural network algorithm, but accuracy rate is not high enough, and unstable training results are the main problems. In this paper, changed trend of learning rate is used to overcome them. From the experimental results it can be known that using these strategies, the Segmented Learning Algorithm is better than the other two methods in the training process of BP neural network.

References

[1]. Yan Zhang, Geng Guohua, Zhou Quanming, Wang Xiaofeng. A feature extraction algorithm based on the music melody outline. Computer Engineering and Applications 2008;44(18):12-14

[2]. Li Jian. Research on Application of Music Classification Based on Neural Network. Computer Simulation 2010; 27(11):168-171

[3]. Kurita, T.A method to determine then number of hidden units of three-layered neural networks by information criteria[J]. The Transactions of the Institute of Electronics, Information and Communication Engineers(in Japanese), J73-D-I I, 1872-1878

[4]. Riedmiller M, Braun H.A Direct Adaptive Method for Faster Backpropagation Learning:the RPROP Algorithm. Processings of the IEEE international Conference on Neural Networks(ICNN) San Fraucisco, 1-993:586-591

[5]. Wang Xiaochuan, Shi Feng, Yu Lei, LiYang. MATLAB-Analysis of 43 cases of neural network. Beijing:Beijing University Press. 2011:1-10

[6]. R.A.Jacobs.Increased rates of convergence through learning rate adaptation. Neural Networks. 1988,vol.1,no.4,pp.295-308

[7]. Zhaoqi Bian, Xuegong Zhang. Pattern recognition. Beijing:Tsinghua University Press. 2000:185-197

[8]. Paul Scott. Music Classification using Neural Networks. Manuscript Class ee373a, Stanford, 2001

[9]. Q.Hong&H.Zhenya,Variable Step BP Algorithm Which Prunes Away Redundaant Connections Dynamically, IJCNN, 1992, No.2, 441-445, Beijing

[10]. Charlambous.Conjugate gradient algorithm for effient training of artificial neural network ,IEE Proceedings,1992,vol.139,no.3,pp.301-310

[11]. V.V.Phansalkar & P.S Sastry, Analysis of The Back-Propagation Algorithm With Momentum ,IEEE Trans ,on NN,1994,Vol.5,No.3,505-506

Wei Gao received the B.Eng degree in Computer Science and Technology from Shenyang University of Chemical Technology and M.Eng degree and ph.D.Eng degree in College of Information Science and Engineering of Northeastern University. Her current research interests on Pattern Recognition and Intelligent Information Processing.

Yadong Yang received the B.Eng degree in Computer Science and Technology from Shenyang University of Chemical Technology. He is currently working toward the M.Eng degree at the same school. He is researching on Computer Application Technology.

Yi Li received the B.Eng degree in Computer Science and Technology from Shenyang University of Chemical Technology. He is currently working toward the M.Eng degree at the same school. He is researching on Virtual Machine Communication Security.