# Music emotion recognition algorithm based on BP neural network

*Xu* Liu*, and *Pingxiao* Ge

Animation Art College, Jilin Animation Institute, Changchun, Jilin, China

**Abstract.** Music plays a very important role in animation production. Because it could better express the emotion of the character, this paper uses BP neural network to identify the music emotion. This paper first introduced the structure of BP neural network. Then, the parameters and structure of the network were designed according to the category of music emotion. Finally, a three-layer BP neural network with 5 input nodes, 13 hidden layer nodes and 4 output nodes was constructed and applied to music emotion recognition. The recognition accuracy was 85.02%, which basically met the requirements of music emotion recognition and achieves the expected effect.

**Keywords:** BP neural network, Music emotion, Animation production.

## 1 Introduction

As an art of expressing human emotions, it uses computers to combine the basic elements of music, and shows a rich emotional world. In animated movies, music can be used to render the emotions and actions of character. Using computer to explore the mapping relationship between musical features and character action space, which actually built a model of emotion recognition in music. Computers could not directly capture the emotional types of music, so it can only be represented by the characteristic information of the music. Only when we fully understand the musical characteristics of music emotion, can we more accurately grasp the emotional connotation of music and drive the role.

At present, the academic circle has studied on music recognition and music emotion for decades. Sun Hui[1] proposed a method of combining the optimal kernel function to form a composite kernel function for genre classification. By selecting the most influential feature combination for classification, the classification results have been significantly improved. Li Hongwei et al.[2] proposed that dynamic brain network was used to study long-term music emotion. Gaussian mixture model and Bayesian classifier were used to classify music emotion, and its accuracy was higher than that of traditional classifier or classification strategy. Liu Wanjun et al.[3] used deep convolution neural network for music genre recognition, and the recognition accuracy was 4% ~ 20% higher than other machine learning models. Liu Mingxing [4] used BP neural network to establish an automatic music

*Corresponding author: 326785495@qq.com

emotion recognition model, extracted Mel cepstrum coefficient MFCC, average zero crossing rate, short-term energy, beat number and other music emotion characteristic parameters to identify music emotion.

In this paper, the advantages of BP neural network was used for music emotion recognition. In the process of recognition, a BP neural network model matching with music recognition was constructed. Finally, this model was used for experimental analysis.

## 2 BP neural network

BP neural network is a kind of back propagation network, which can solve the problem of network performance change when a single weight changes. The self-organization and self-learning ability of BP neural network makes it possible to learn from large samples. Then the nonlinear relationship between emotion and cause is established, and the amount of computation consumed is very small, so it has good practical value in the process of emotion recognition.

BP neural network has three layers of neurons, as shown in Figure 1. It mainly includes input layer, hidden layer and output layer.

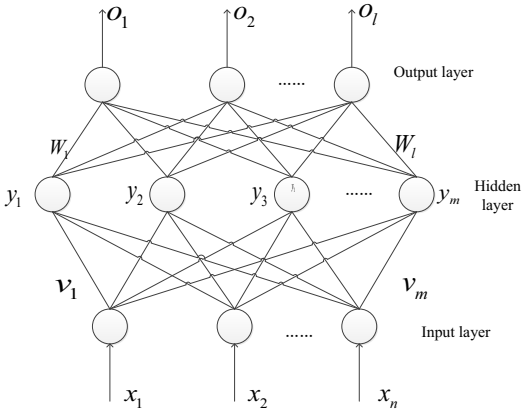


**Fig. 1.** Structure of BP neural network.

In Figure 1, the input vector was $X=(x_1,x_2,...,x_n)^T$. The output vector of the hidden layer was $Y=(y_1,y_2,...,y_m)^T$. The output vector was $O=(o_1,o_2,...,o_m)^T$. The weight vectors between input layer and hidden layer, hidden layer and output layer were $V=(v_1,v_2,...,v_m)^T$ and $W=(w_1,w_2,...,w_l)^T$ respectively, where $v_j$ and $w_k$ are the weight vectors corresponding to the $j$-th neuron in the hidden layer and the $k$-th neuron in the output layer respectively.

$$\text{Input: } net_j=\sum_{i=0}^{n} v_{ij}x_i,\ j=1,2,...,m \tag{1}$$

$$\text{Output: } y_j=f(net_j),\ j=1,2,...,m \tag{2}$$

The input and output of the output layer are as follows:

$$\text{Input: } net_k=\sum_{i=0}^{n} w_{jk}y_j,\ k=1,2,...,l \tag{3}$$

$$\text{Output: } o_k = f(net_k),\ k = 1,2,...,l \tag{4}$$

where $f(\cdot)$ is the excitation function and Sigmoid function is used in this paper

$$f(x) = \frac{1}{1+e^{-x}} \tag{5}$$

In practical application, the output value is not consistent with the expected value. There is a certain deviation, so that the expected vector is $d = (d_1, d_2, ..., x_l)^T$. Then the error $E$ of them is:

$$E = \frac{1}{2}\sum_{k=1}^{l}(d_k - o_k)^2 \tag{6}$$

By substituting formula (4) into formula (6), it is obtained that:

$$E = \frac{1}{2}\sum_{k=1}^{l}(d_k - f(net_k))^2 = \frac{1}{2}\sum_{k=1}^{l}(d_k - f(\sum_{i=0}^{n} w_{jk} y_j))^2 \tag{7}$$

Then, by substituting equation (2) into equation (7), we can get the following results:

$$E = \frac{1}{2}\sum_{k=1}^{l}(d_k - f(\sum_{i=0}^{n} w_{jk} f(net_j)))^2 = \frac{1}{2}\sum_{k=1}^{l}(d_k - f(\sum_{i=0}^{n} w_{jk} f(\sum_{i=0}^{n} v_{ij} x_i)))^2 \tag{8}$$

It can be seen that the output error is related to the weight vector of each layer, and the error can be reduced by adjusting $w_{jk}$ and $v_{ij}$. therefore, in the process of adjustment, the change of the weight should be increased with the increase of the negative gradient of the error, that is:

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{kj}},\ \ j = 0,2,...,m; k = 1,2,...,l \tag{9}$$

$$\Delta v_{ij} = -\eta \frac{\partial E}{\partial v_{kj}},\ \ j = 0,2,...,n; k = 1,2,...,m \tag{10}$$

$\eta$ is (0,1), which reflects the learning rate through the change of negative gradient.
By further derivation of equation (9), it can be concluded that:

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{kj}} = -\eta \frac{\partial E}{\partial net_k}\frac{\partial net_k}{\partial w_{kj}} = \eta(d_k - o_k)o_k(1-o_k)y_j \tag{11}$$

Let $\delta_k^o = (d_k - o_k)o_k - (1-o_k)$, then equation (11) can be rewritten as:

$$\Delta w_{jk} = \eta \delta_k^o y_j \tag{12}$$

In the same way, we can find that:

$$\Delta v_{ij} = -\eta \frac{\partial E}{\partial v_{kj}} = -\eta \frac{\partial E}{\partial net_j}\frac{\partial net_j}{\partial v_{kj}} = \eta\left[\sum_{k=1}^{l}(d_k - o_k)o_k(1-o_k)y_j\right]y_j(1-y_j)x_i = \eta\left(\sum_{k=1}^{l}\delta_k^o y_j\right)y_j(1-y_j)x_i \tag{13}$$

If $\delta_j^y = \left( \sum_{k=1}^{l} \delta_k^o y_j \right) y_j (1 - y_j)$ is the output error of the hidden layer signal, then we can get:

$$\Delta v_{ij} = \eta \delta_j^y x_i \tag{14}$$

It can be seen from the above derivation that the error of output layer is related to the actual output and expected output of output layer, and the output error of hidden layer is related to the signal transmitted by output layer. So it can be said that BP neural network is to pass the weight through the back propagation of error to realize the training of neural network. The positive and negative propagation of the signal can be shown in the Figure 2.

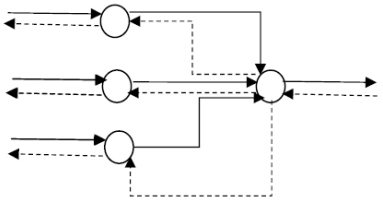

**Fig. 2.** Propagation of positive and negative signals of BP neural network.

## 3 Design of BP neural network model

When applying BP neural network to emotion recognition, it is necessary to design an appropriate network according to the actual situation, that is, to design the network parameters and network structure. The former mainly includes the design of input layer, output layer feature vectors, learning samples and initial weights. The latter includes the number of network layers and the number of nodes in different layers.

### 3.1 Parameter design of BP neural network

#### *3.1.1 Determine the input vector*

The input vector number is appropriate, which can effectively reflect the emotional characteristics of the robot. If the number of input vectors selected is too few, the feature vectors can not effectively recognize and classify the robot's emotion, while too many vectors will reduce the convergence of the network. There were five input layers selected in this paper, which were pitch, length, timbre, speed and strength. The input value range of the node was [0,1].

#### *3.1.2 Determine the output vector*

The output vector was the emotion type of music. There were four nodes in the output layer: relaxed, excited, sad and angry emotion. However, the value range of output was 0 or 1, which indicated whether the emotion was activated.

#### *3.1.3 Identify learning samples*

The collection of learning samples is very important for the detection results of BP neural network. When selecting learning samples, we should follow the principle that the samples

should be representative, extensive and compact. In this paper, 120 groups of data are selected as learning samples.

#### *3.1.4 Design initial weight*

BP neural network is a nonlinear system. Local minimum, convergence speed and training time should be considered when selecting initial weights. In general, the output value is zero by adjusting the initial weight, so the error is minimized by adjusting the learning function. Generally, the range of initial weight is (-1, 1).

### 3.2 Structure design of BP neural network

#### *3.2.1 Design of network layers*

A lot of practice had proved that the three-layer BP neural network was enough to achieve any nonlinear mapping, which could also meet the accuracy requirements, so this paper selected the three-layer BP neural network.

#### *3.2.2 Design of input node*

According to the input vector of input nodes, the number of input nodes used in this paper were 5.

#### *3.3.3 Design of output node*

The number of output nodes of the network were the type of emotion, so it could be seen that the number of output nodes of the BP neural network selected in this paper was 4.

#### *3.3.4 Design of hidden layer node*

The number of hidden layer nodes directly affected the performance of neural network system. If the number of hidden layer nodes was too few, the fault tolerance and recognition ability of the system would be poor. On the contrary, if the number of hidden layer nodes was too many would make the training time of the system too long and reduce the generalization ability.

According to the empirical knowledge, the following formula can be used to calculate the number of hidden layer nodes.

$$m \geq \sqrt{n+l} + \beta \quad or \quad m \geq \sqrt{nl} \tag{15}$$

where $m$ is the number of nodes in the hidden layer, $n$ and $l$ are the number of input and output nodes respectively, $\beta \in [1,10]$. In this paper, the number of neurons in input layer and output layer is 5 and 4 respectively, so the number of nodes in hidden layer is 5-14.

## 4 Analysis of experimental results of music recognition based on BP neural network

In order to fully express the emotion of music, this paper used the four most representative emotion types of Thayer emotion model, which were excited, relaxed, sad and angry.

In this paper, when using BP neural network for music recognition, the transfer function was sigmoid function, the number of hidden layer nodes was 13, and the training function was *trainrp*. From MIDI music library, 120 songs with different styles and covering all emotions were selected for recognition. The recognition results were shown in Table 1.

**Table 1.** Results of music recognition based on BP neural network.

| Music length (s) | Number of samples | Recognition rate | Average recognition rate |
|---|---|---|---|
| 0-50 | 32 | 88.9 | 85.02 |
| 50-100 | 30 | 87.3 | |
| 100-150 | 26 | 85.6 | |
| 150-Over 200 | 22 | 82.1 | |

The experimental results showed that the accuracy of the long automatic recognition model of music emotion was obviously reduced. The accuracy rate of the method proposed in this paper reached 85.02%, which basically met the requirements of music recognition and achieves the expected effect.

## 5 Conclusion

This paper introduced the network structure and implementation principle of BP neural network in detail, and designed the parameters and network of BP neural network applied to music emotion recognition in detail. Finally, this paper used 5 nodes in the input layer, 4 nodes in the output layer, 13 nodes in the hidden layer, and the training function was *trainrp* to recognize the emotion of the collected music. The final recognition accuracy was 85.02%. The music emotion will be obtained later to drive the change of the character action, which is convenient for the animation to be generated quickly.

Supported by key projects of social science research of Jilin Animation Institute (No. 2019006).

## References

1. Li Hongwei, Li Haifeng, Ma Lin. long term music emotion research based on dynamic brain network[J]. Journal of Fudan University (Natural Science), 2020, v.59 (03): 82-89
2. Sun Hui, Xu Jieping, Liu Binbin. Music genre classification based on multi-core learning support vector machine[J]. Journal of Computer Applications, 2015, 35 (6): 1753-1753
3. Liu Wanjun, Meng Renjie, Qu Haicheng, Liu Lamei. Research on music genre recognition based on enhanced alexnet[J]. CAAI transactions on intelligent systems, 2020, V. 15; No.84(04):124-131.
4. Liu Mingxing. Music classification model based on BP neural network[J]. Modern electronics technique, 2018, 41 (05): 144-147