Neurocomputing 105 (2013) 100–106

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Web music emotion recognition based on higher effective gene expression programming

Kejun Zhang *, Shouqian Sun

College of Computer Science, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Available online 8 October 2012

Keywords:
Music emotion recognition
Evolutionary algorithm
Gene expression programming
Support vector machine
Music information retrieval

ABSTRACT

In the study, we present a higher effective algorithm, called revised gene expression programming (RGEP), to construct the model for music emotion recognition. Our main contributions are as follows: firstly, we describe the basic mechanisms of music emotion recognition and introduce gene expression programming (GEP) to deal with the model construction for music emotion recognition. Secondly, we present RGEP based on backward-chaining evolutionary algorithm and use GEP, RGEP, and SVM to construct the models for music emotion recognition separately, the results show that the models obtained by SVM, GEP, and RGEP are satisfactory and well confirm the experimental values. Finally, we report the comparison of these models, and we find that the model obtained by RGEP outperforms classification accuracy of the model by GEP and takes almost 15% less processing time of GEP and even half processing time of SVM, which offers a new efficient way for solving music emotion recognition problems; moreover, because processing time is essential for the problem of large scale music information retrieval, therefore, RGEP might prompt the development of the music information retrieval technology.

1. Introduction

Music emotion recognition, as one of the main research field of music information retrieval, has been researched for many years [1], and there are many algorithms, frameworks and applications have been proposed in the literature [2,3], recently, Yang et al. [4] proposed a harmonizing hierarchical manifolds for cross-media retrieval and presents a new framework for multimedia content analysis and retrieval based on semi-supervised ranking and relevance feedback[5]. Obtaining a good-quality model for music emotion recognition depends on many factors, like the selection of statistical methods and feature extraction, for the later, recently, Ma et al. [6] proposed a novel feature selection method and applied it to automatic image annotation. In the study, we will mainly concern about the statistical methods, although support vector machine (SVM) has been chosen as one of the best statistical methods [7,8], it can be time-consuming. Therefore, there remains a need for an efficient method for improving the performance of music emotion recognition.

Studies show that artificial intelligence ways are efficient to find out acceptable model for music emotion recognition [9–11]. Recently, gene expression programming (GEP) [11], as the natural development of genetic algorithms [12,13] and genetic programming [14], has been used to deal with problem of music information retrieval, especially, Yang et al. [15] used GEP to solve the problem of music emotion recognition (for MIDI music clips), and the literature suggested that GEP can improve efficiency of model for music emotion recognition.

However, over the years, many criticisms of GEP have been as follows:

1. Lack of effective selection method
   Roulette-wheel sampling with elitism [16] has been used in GEP for many years, to the author's knowledge, little attention has been devoted to the selection method of GEP [11,15].
2. Processing time
   Until recently, there is little information available in the literature about how to decrease the processing time of GEP [11,15].

In the study, we address all of these issues and present a higher effective algorithm, called revised gene expression programming (RGEP), to construct the model for music emotion recognition. From the experimental results, we find that the model obtained by RGEP outperforms classification accuracy of the model by GEP and takes less processing. Moreover, a fuzzy exploring method support for "emotion vector search" will be provided too.

The paper is organized as follows: In Section 2, we give a brief introduction to the methods we used in the study: GEP, RGEP, and SVM. After a detailed introduction of the experiments in Section 3,

* Corresponding author.
E-mail address: channy@zju.edu.cn (K. Zhang).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.06.041

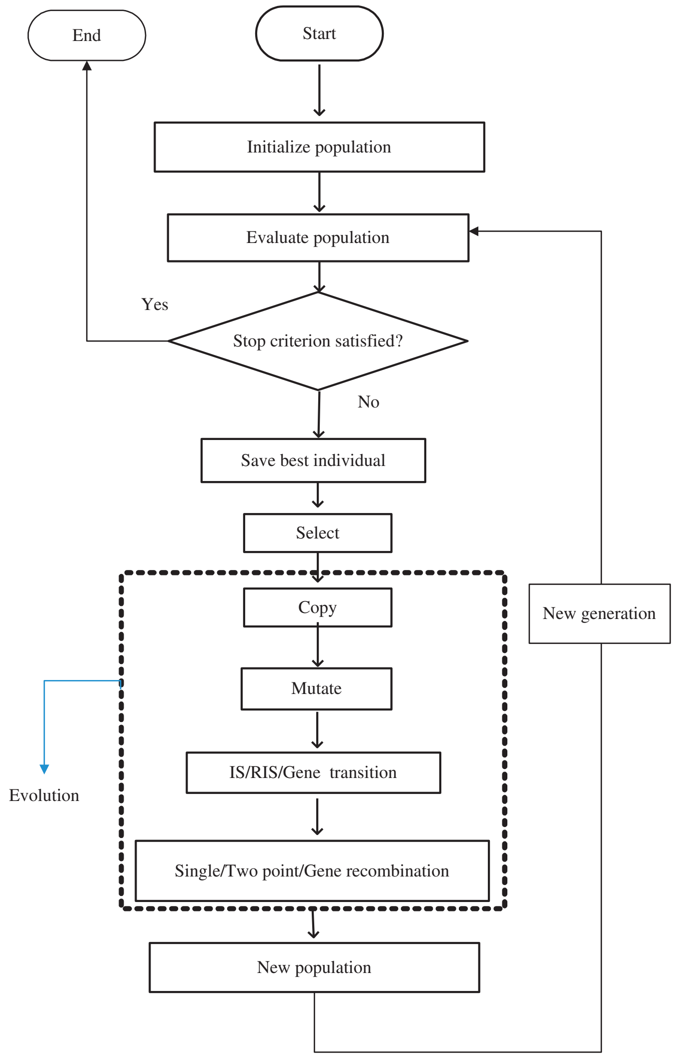

Fig. 1. Gene expression programming.

we present the experimental results and discussions in Section 4. Finally, we summarize the key conclusions of the study in Section 5.

2. Methods

2.1. Gene expression programming

GEP uses the same kind of diagram representation of genetic programming, but the entities evolved by expression tree are the expression of a genome and individuals are often copied into the next generation based on their fitness, as determined by roulette-wheel sampling with elitism, which guarantees the cloning of the best individual to the next generation [11]. Fig. 1 shows the flowchart of GEP.

As you can see from Fig. 1, GEP starts from an initial population with many genes (individuals), after generations of evolution, the best individual will be selected and its decoding process can be expressed by Fig. 2 [11]. In Fig. 2 sin represents function sin() and terminal x_1–x_{10} represent the variables; the alleles represent the position in the genes (individuals), as you can see from Fig. 2, according to GEP grammar [11], the individual will be expressed as Expression Tree and the Expression Tree can be easily decoded as a math equation. Detailed description of GEP will be found in [11].

Here, we take an example to explain how GEP works [11]:

Step 1: INITIALIZATION

We set the individuals with a length of 17 and the function set: $F=\{‘+’,‘-’,‘\times’,‘/’,‘S’,‘C’,‘Q’\}$, here, ($S$ represents function sin(), C represents function cos(), Q represents function sqrt(). Then, we encode all the variables (here, simply, we just use the 7 features with variables x_1–x_7) and functions into one

Allele	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Value	C	/	+		-	+		+	x_6	x_3	x_7	x_1	x_4	x_5	x_1	x_6	x_4

Fig. 3. Individual with length of 17.

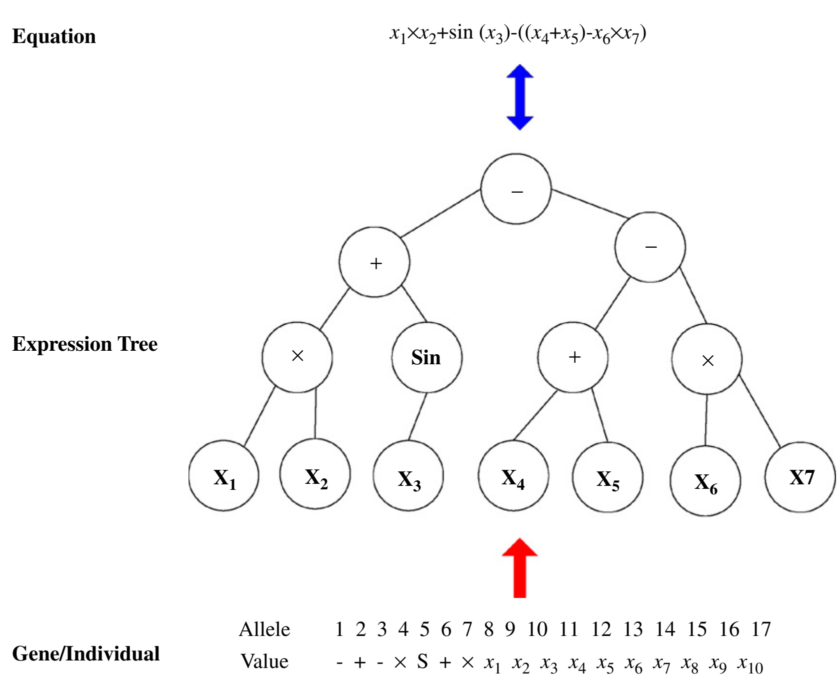

Fig. 2. Decoding process of individual with one gene.

chromosome and randomly initialize the population. For example, one of the individuals may be like Fig. 3.

With this individual, we can easily decode it as an Eq. (1):

$$F(X)=\cos\left(\frac{x_6-x_3+x_7+x_7}{x_4x_5(x_1+x_6)}\right) \tag{1}$$

Here, $F(X)$ is the value of relevant emotions in our study.

Step 2: OPERATION

For each generation, we implement GEP operators. For the chromosome in Fig. 3, if we implement mutation operator in allele 1 and change the value to square function, then we will get an Eq. (2):

$$F(X)=\sqrt{\frac{x_6-x_3+x_7+x_1}{x_4x_5(x_1+x_6)}} \tag{2}$$

Please see [11] for details about how other operators implement.

Step 3: FITNESS CALCULATION

We use the Eq. (3) to calculate the fitness:

$$Fitness=1000/(MSE+1) \tag{3}$$

Here, MSE is the mean square error, we will describe it later.

Step 4: STOP CRITERION

If the stop criterion satisfied, stop, else, go to step 2. In general, stop criterion is related to fitness.

2.2. Revised gene expression programming

In this section, we design and implement a revised gene expression programming based on back-chaining evolutionary algorithm [17], (see Algorithm 1) and the relevant time and space complexity is also provided.

1: Let $\boldsymbol{r}$ be an individual in the population at generation $\boldsymbol{G}$
2: Choose an operator to apply to generate $\boldsymbol{r}$
3: Do tournaments to select the parents:
  $\{\boldsymbol{s_1}, \boldsymbol{s_2} \ldots\}$=individuals in generation $\boldsymbol{G}-\boldsymbol{1}$ involved in the tournaments
4: Do recursion using each unknown $\boldsymbol{s_j}$ as a sub goal recursion terminates at generation 0 or when the individual is known (i.e. has been evaluated before)
5: Repeat for all individuals of interest in generation $\boldsymbol{G}$

Algorithm 1. Backward-chaining evolutionary algorithm

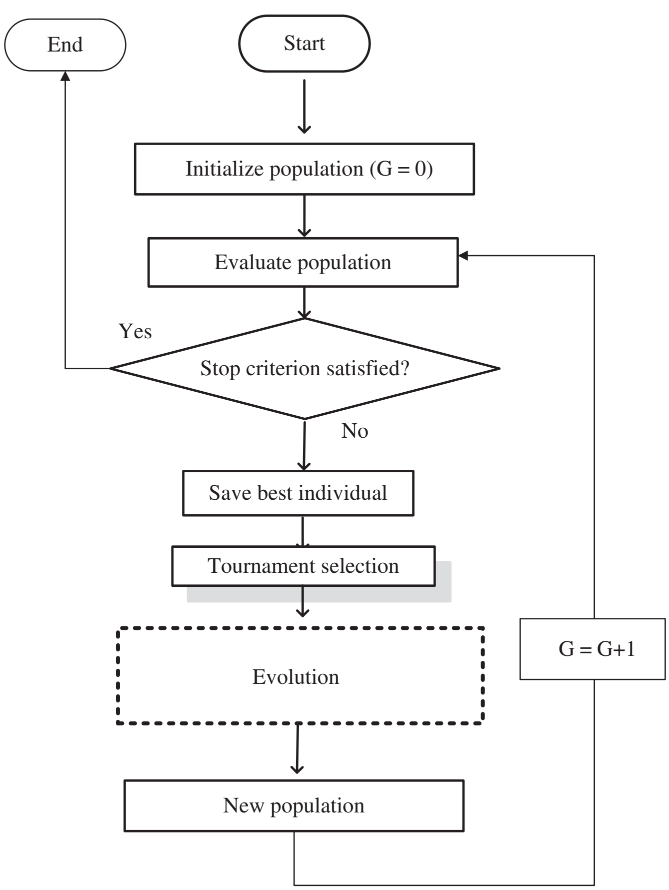

Fig. 4. GEP based on tournament selection.

Then, as you can see from Fig. 4, we use the tournament selection[18] instead of roulette-wheel sampling with elitism, so as we can employ backward-chaining evolutionary algorithm, which obtained time savings by avoiding the creation and evaluation of individuals not sampled by selection (and their unnecessary ancestor) [17].

Finally, following the ideas in GEP based on tournament selection and backward-chaining evolutionary algorithm, we propose RGEP (see Algorithm 2).

```
Procedure run RGEP (G, P)
G means the generation, I means individual, and P means the population size
    begin
        initialize population and the fitness,
        create arrays solved
        for all individuals I of interest in generation G do
            evolve_back (I, G)
        end for
        return all I of interest
    end
Procedure evolve_back (indiv, gen)
    if solved [indiv] [gen] then
        return
    end if
    if random_float () < recombination_rate then
        parent1 = tournament (gen − 1)
        parent2 = tournament (gen − 1)
        population [gen] [indiv] = recombination (parent1, parent2)
    end if
    if random_float () < transposition _rate then
        parent1 = tournament (gen − 1)
        parent2 = tournament (gen − 1)
        population [gen] [indiv] = transposition (parent1, parent2)
    end if
    if random_float () < mutation_rate then
        parent = tournament (gen − 1)
        population [gen] [indiv] = mutation (parent)
    end if
Procedure tournament (gen)
    fbest = 0
    best = undefined
    for tournament_size times do
        candidate = random inter 1…P
        evolve_back (candidate, gen)
        if Fitness [gen] [candidate] > fbest then
            fbest = Fitness [gen] [candidate]
            best = candidate
```

9: **end if**
10: **end for**
11: **return** Population [***gen***] [***best***]

Algorithm 2. Revised gene expression programming

RGEP is based on changing the order of various operations in a GEP, which offers a combination of fast convergence, increased efficiency in terms of fitness evaluations, complete statistical equivalence to a standard GEP and broad applicability [17], however, it requires memorizing choices and individuals over multiple generations. According to the work of [17], in our experiments, we get the following results:

1. For space complexity
   In the worst possible case (where all programs are constructed and evaluated, so, the number of individuals actually created and evaluated during the run of RGEP are same as the number of individuals during the run of GEP), if the maximum size of the individuals in each generation throughout the run is 100 bytes, we just need Ω of memory space, Ω equals to:
   $$\Omega = (G_{\max}+1) \times P \times 100 \tag{4}$$
   where $G_{\max}$ is the maximum generation, P is the size of population.
   For example, if we run a population of $P=100$ individuals with maximum generation of 600 for 100 times, without memory optimization, RGEP requires only around $101 \times 100 \times 600 \times 100$ bytes, 600 MB of memory to run, which is readily available in most modern personal computers.
2. For time complexity
   For GEP we have
   $$T_{GEP} = M \times (G+1) \times P \times S_{avg}^{GEP} \tag{5}$$
   And for RGEP
   $$T_{RGEP} = M \times E_{RGEP} \times S_{avg}^{RGEP} \tag{6}$$
   Obviously, the time saving provided by RGEP can be expressed as an Eq. (7):
   $$T_{save} = T_{GEP} - T_{RGEP} = M \times \left[(G+1) \times P \times S_{avg}^{GEP} - E_{RGEP} \times S_{avg}^{RGEP}\right] \tag{7}$$
   where G is the current number of generation, E_{RGEP} is the number of individuals actually created and evaluated during the RGEP runs. $M$is the amount of training sets, S_{avg}^{GEP} and S_{avg}^{RGEP} is the average size of the (created) individuals during the GEP runs and RGEP runs respectively.

Obviously, E_{RGEP} is less than $(G+1) \times P$; and considering bloat[17], $S_{avg}^{RGEP}$is much smaller than S_{avg}^{GEP}. Therefore, compared to GEP, much time will be saved in RGEP; moreover, for a bloating population the parsimony of RGEP in terms of fitness evaluations is compounded with its parsimony in terms of program sizes to produce even more impressive savings [16].

2.3. Support vector machine

SVM is a type of machine learning method and a kernel-induced feature space function is used for the mapping of objects onto target values [7]. In SVM, a kernel-induced feature space with function $k(x,x_i)$ is used for the mapping of objects onto target values. Thus a non-linear feature mapping will allow the treatment of non-linear problems in a linear space. The prediction or approximation function used by a basic SVM is an Eq. (8):

$$f(x) = \sum_{i=1}^{l} a_i K(x,x_i) + b \tag{8}$$

where a_i is some real value, x_i is a feature vector corresponding to a training object, and $k(x_i,x)$ is a kernel function. The components of the vector **a** and the constant b represent the hypothesis and are optimized during training. $k(x_i,x)$ is a kernel function, which value is equal to the inner product of two vectors x and x_i in the feature space $\varphi(x)$ and $\varphi(x_i)$. In this study we use a Gaussian radial basis function as the kernel function.

$$K(u,v) = \exp\left(-\gamma^* |u-v|^2\right) \tag{9}$$

where γ is constant, u and v are two independent variables, the generalization of SVM can be controlled by adjusting the shape of Gaussian with γ. Details about SVM, please see [7].

3. Experiments

3.1. Database

A total of 726 main rhythm parts (music clips with 30 s) of relevant popular songs with MP3 format was selected in the study. According to [19], we use MARSYAS [20] to retrieve 30 features from each music clip, which includes 19 of timbral texture features (means and variances of spectral centroid, roll-off, flux, zerocrossings over the texture window (8), low energy (1), and means and variances of the first five MFCC coefficients over the texture window (excluding the coefficient corresponding to the DC component)) 6 of rhythmic content features, 5 of pitch content features [19].

As reported in [21], supplementary training in music schools or not is negligible to determine the musical ability (emotion experience) of human beings; therefore, we recruit 80 students to annotate the music clips and do not classify the students with the background. To make it more objective, the music clips we used are unknown to the students. According to Hevner model [22], we conducted a study on adjective selection for music emotion space [23,24], and 8 adjectives selected for labeling music emotion: Vigorous, Dignified, Sad, Dreaming, Soothing, Graceful, Joyous, and Exciting. Then, students need to annotate each music clip with these 8 emotion labels, set a value from 0 to 1 to represent the similarity for each emotion (the value close to 1 means very similar), for example, a song can be annotated as:

$$\mathbf{Music}_{Emotions} = [0.3,0.2,0.2,0.1,0.3,0.2,0.6,\mathbf{0.9}] \tag{10}$$

where, the values represent the similarity to Vigorous, Dignified, Sad, Dreaming, Soothing, Graceful, Joyous, and Exciting respectively. For example, the value of "Sad" is 0.2, it means "not sad"; the value of "Exciting" is 0.9, it means "very exciting". In the study, the emotion with maximum value (bold one) will be defined as the song's main emotion, so the example song's main emotion is "exciting".

After annotation, most of the students annotated all the music clips, if a music clip has been annotated and the main emotion got less than 70% of the total annotation, we consider the music clip as hard distinguishable one and delete it from the database. In sum, we get 46,648 annotations: each of the 686 songs is annotated by 68 participants. We choose the average value as the final value to construct the whole database. Fig. 5 shows the procedure of database construction for music emotion recognition.

3.2. Evaluation

To evaluate the system, two classical equations are used: Mean square error (MSE) of program (individual) i and correlation

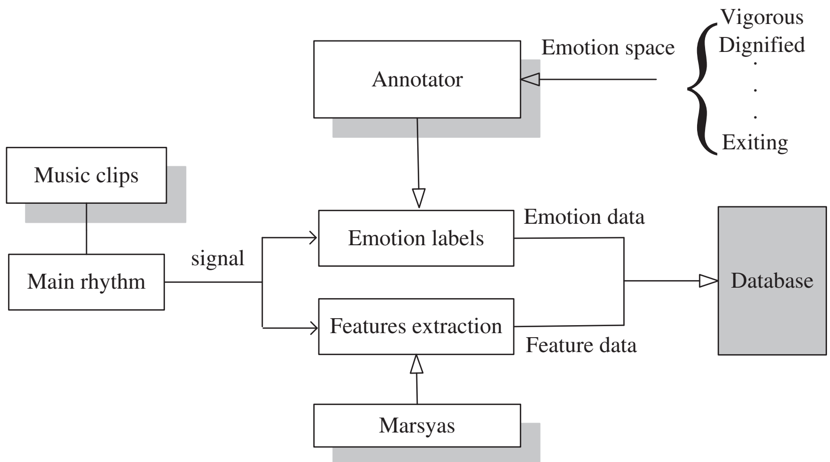

Fig. 5. Procedure of database construction for music emotion recognition.

coefficient (CC).

$$MSE_i = \frac{1}{n}\sum_{j=1}^{n}\left(F_{(ij)}-Q_j\right)^2 \tag{11}$$

where $F_{(ij)}$ is the value predicted by the individual program i for the fitness case j (out of n fitness cases) and Q_j is the target value for the fitness case j.

$$CC = \frac{Cov(F,Q)}{\sigma_f \sigma_q} \tag{12}$$

where $Cov(F,Q)$ is the covariance of the target and model outputs, σ_f and σ_q are the corresponding standard deviations.

3.3. Model for music emotion recognition

Music emotion recognition is a process of modeling between music features space and emotion space (see Fig. 6). In Fig. 6, ***X*** means music features space; ***Y*** means emotion space. ***F*(*X*)** is the expected model.

Detailed information about modeling for music emotion recognition by SVM (how to find the best ***F*(*X*)**), please see [7].

Fig. 7 shows the model for Music emotion recognition based on GEP and RGEP:

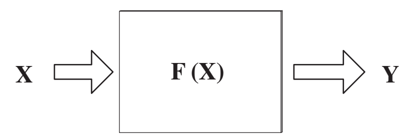

Fig. 6. Model for music emotion recognition.

4. Results and discussions

We selected 600 music clips as the training set and the left 86 music clips as the testing set, and we will mainly concern about the main emotion recognition. Therefore, three steps will be made in achieving the comparison results:

Step 1: Train the classifier for 100 times separately based on training set and compare the relevant processing time in GEP, RGEP and SVM;
Step 2: Conduct the experiments for many times to find the best model;
Step 3: Test the best model on testing set.

Details will be described in the following sections.

4.1. GEP and RGEP

In the study, all the parameters and the concepts of GEP and RGEP are used with reference to literature [11]. Firstly, in the training step, we divide the training set into 10 parts and do 10-fold cross validation test. The experiments show that GEP and RGEP both get MSE below 0.020. Then, taking MSE=0.020 as the stop criterion, we conduct the experiment for the whole training datasets by GEP and RGEP for 100 times (in same condition with same parameters (see Table 1), other parameters are the same, please see [11]), each run stopped only if the stop criterion satisfied or beyond 10 min of the maximum run time limit.

Secondly, we conduct the experiments with maximum generation G=600 (without stop criterion) on the whole training set for many times by GEP and RGEP, and save the best models with a minimum MSE of GEP and RGEP respectively (see Table 2).

Finally, test the best model on testing set separately and get the compare results in Table 3.

4.2. SVM

In the study, the method was implemented based on WEKA and LibSVM [25,26]. Firstly, in the training step, we need to find the optimal γ, ε, C. According to the generalization ability of the model based on the Leave One Out Cross Validation for the training set, we get the best parameters of γ, ε and C are 0.006, 0.015, and 68 respectively. With these parameters, taking MSE=0.020 as the stop criterion, we conduct the experiments on the whole training set for 100 times (same as GEP and RGEP).

Secondly, we conduct the experiments with time limits of 10 min (without stop criterion) on the whole training set for many times by SVM, and save the best model with minimum MSE (see Table 2).

Finally, test the best model on testing set and get the compare results in Table 3.

4.3. Discussions

In the study, all the experiments are conducted in Win7 with Intel core 2 Dua 2.53 G and 4 G memory. As you can see from Table 2, RGEP outperforms classification accuracy of GEP and takes least processing time (84.62% of GEP and almost half of SVM).

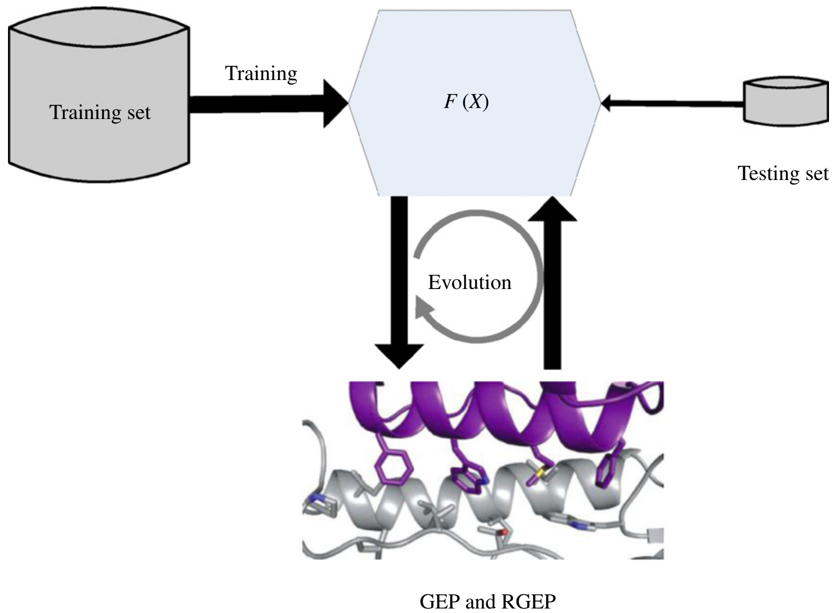

Fig. 7. Model for music emotion recognition based on GEP and RGEP.

Table 1
Parameters used in GEP and RGEP.

Parameters	Value
Size of population	100
Function set	$F=[+, -, \times, \div, S, C, Q]$
Terminal	$T=[x_1\text{–}x_{30}]$
Numbers of genes, gene head size	5, 15
Mutation rate	0.044
One/two point combination	0.2
Gene recombination	0.05
IS/RIS/Gene transposition rate	0.05

Table 2
Comparison result of music main emotion recognition (Training set).

Algorithm	*CC*	*MSE*	100 runs time (s)
SVM	0.782	0.0072	31,782
GEP	0.6425	0.0193	15,892
RGEP	0.7602	0.0157	13,448

Table 3
Music main emotion recognition results (Testing set).

Algorithm	Success rate (%)			
	$\Phi=0.005$	$\Phi=0.010$	$\Phi=0.015$	$\Phi=0.020$
SVM	87.21	87.21	87.21	87.21
GEP	83.72	84.88	84.88	84.88
RGEP	81.40	87.21	87.21	87.21

We can see from Table 3, the best recognition results are 87.21%, 84.88%, and 87.21% in SVM, GEP and RGEP respectively. And the best value for Φ might between 0.010 and 0.015 for them.

Here, a threshold δ is used, δ can be defined as

$$\delta=|F-Q| \tag{13}$$

F is the predicted value and Q is the relevant experimental value. Then we define a thresholdΦ, if $\delta<\Phi$, we consider it as success testing.

To be mentioned, in recent work, we add more function sets to RGEP, the classification accuracy is extremely improved, however, the processing time increased too. So, the further research should pay more attention to best function sets finding in terms of processing time.

Moreover, like [24], since the RGEP system takes the music clips as an emotion vector; therefore, a song with 8 emotion labels: Vigorous, Dignified, Sad, Dreaming, Soothing, Graceful, Joyous, and Exciting, could be calculated like Eq. (14):

$$\mathbf{Music}_{Emotions}=[F(x)_1,F(x)_2,F(x)_3,F(x)_4,F(x)_5,F(x)_6,F(x)_7,F(x)_8] \tag{14}$$

where $F(x)_1$–$F(x)_8$ represent the value of Vigorous, Dignified, Sad, Dreaming, Soothing, Graceful, Joyous, and Exciting.

Consequently, we can get some fuzzy exploring by using this system. For example, for a song with emotion vector:

$$\mathbf{Music}_{Emotions}=[0.3,0.2,0.2,0.1,0.3,0.2,\mathbf{0.6},\mathbf{0.9}] \tag{15}$$

We can explore this song by key words “very exciting & joyous” (bold one) in the web. It’s a very good and effective way to search for songs.

5. Conclusion

In the study, we present a revised gene expression programming based on backward-chaining evolutionary algorithm to construct the model for music emotion recognition. From the experimental results, we find that the model obtained by RGEP outperforms classification accuracy of the model by GEP and takes almost 15% less processing of GEP and even half processing time of SVM, which offers a new efficient way for solving music emotion recognition problems; And because processing time is essential for the problem of large scale music information retrieval (MIR), RGEP might prompt the development of the web MIR technology. Moreover, a fuzzy exploring method support for “emotion vector search” is provided too.

Further research should include more detailed evaluation of the system performance for large scale datasets, find an appropriate feature selection method, research on multiple label recognition, and also the parameter Φ might be applied into fitness function later.

Acknowledgments

This study is partly supported by the National Natural Science Foundation of China (61070075, 61004116, 61003147).

References

[1] X. Hu, J.S. Downie, C. Laurier, M. Bay, A. Ehmann, The 2007 MIREX audio mood classification task: lessons learned, in: Proceedings of the 9th International Symposium on Music Information Retrieval (ISMIR), September 2008, Philadelphia, PA.
[2] D. Ververidis, C. Kotropoulos, Fast sequential floating forward selection applied to emotional speech features estimated on DES and SUSAS data collections, in: Proceedings of the XIV European Signal Processing Conference, 2006.
[3] M.M. Ruxanda, B.Y. Chua, A. Nanopoulos, C.S. Jensen, Emotion-based music retrieval on a well-reduced audio feature space, in: Proceedings of the 34th International Conference on Acoustics, Speech, and Signal Processing, April 19–24, 2009, Taipei, Taiwan, 4p, in press.
[4] Yang yi, Yue-Ting Zhuang, Fei Wu, Yun-He Pan, Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval, IEEE Trans. Multimedia 10 (3) (2008) 437–446.
[5] Yang Yi, Nie Feiping, Xu Dong, Luo Jiebo, Zhuang Yueting, Pan Yunhe, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 723–742.
[6] Zhigang Ma, Feiping Nie, Yi Yang, Jasper Uijlings, Nicu Sebe, Web image annotation via subspace-sparsity collaborated feature selection, IEEE Trans. Multimedia, http://dx.doi.org/10.1109/TMM.2012.2187179, in press.
[7] C. Cortes, V. Vapnik, Support vector networks, Mach. Learn. 20 (1995) 273–297.
[8] Jianhua xu, An extended one-versus-rest support vector machine for multi-label classification, Neurocomputing 74 (17) (2011) 3114–3124.
[9] S.-B. Cho, Emotional image and musical information retrieval with interactive genetic algorithm, Proc. IEEE 92 (4) (2004) 702–711.
[10] M.H. Sedaaghi, D. Ververidis, C. Kotropoulos, Improving speech emotion recognition using adaptive genetic algorithms, EURASIP (2007) 2209–2213.
[11] C. Ferreira, Gene expression programming: a new adaptive algorithm for solving problems, Complex Syst. 13 (2003) 87–129.
[12] J.H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.
[13] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley, Boston, Massachusetts, United States, 1989.
[14] J.R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Evolution, MIT Press, Massachusetts, 1992.
[15] C. Yang, S. Sun, K. Zhang, T. Liu, Study on music emotion cognition model based on applying the improved gene expression programming, in: Proceedings of the Second Workshop on Digital Media and its Application in Museum & Heritage, 2007, pp. 344–351.
[16] M. Mitchell, An introduction to genetic algorithms, Complex Adaptive Systems, MIT Press, Cambridge, Massachusetts, United States, 1996.
[17] R. Poli, W.B. Landon, Backward-chaining evolutionary algorithms, Artif. Intell. 170 (2006) 953–982.
[18] T. Back, D.B. Fogel, T. Michalewicz (Eds.), Evolutionary computation Q: Basic Algorithms and Operators, Institute of Physics Publishing, 2000.
[19] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, IEEE Trans. Speech Audio Proc. 10 (5) (2002) 293–302.
[20] G. Tzanetakis, P. Cook, MARSYAS: a framework for audio analysis, Organized Sound, 4, Cambridge University Press, Cambridge, United Kingdom, 2000.
[21] E. Bigand, B. Poulin-Charronnat, Are we experienced listeners? A review of the musical capacities that do not depend on formal musical training, Cognition 100 (2006) 100–130.
[22] K. Hevner, Experimental studies of the elements of expression in music, Am. J. Psychol. 48 (1936) 246–268.
[23] Tao Liu, Research on Music Emotion Recognition Model and Interactive Technology[D], Zhejiang University, Hangzhou, China, 2006.
[24] Bin Zhu, Kejun Zhang, Music emotion recognition system based on improved GA-BP, in: Proceedings of the 2010 International Conference on Computer Design and Applications (ICCDA), vol. 2, 2010, V2-409–V2-412.
[25] I. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition, Morgan Kaufmann Publishers Inc., San Francisco, 2005. Available from: 〈http://www.cs.waikato.ac.nz/ml/weka/〉.
[26] C.-C. Chang, C. Lin, LIBSVM: a library for support vector machines, 2001.

Kejun Zhang received Ph.D. in Computer Science from Zhejiang University, China. He is currently a Post-doctor in the College of Computer Science, Zhejiang University, China. His research interests lies in music information retrieval, bioinformatics, evolutionary computation and machine learning. He is a member of IEEE. He has published many research papers in various reputable journals and conference proceedings.

Shouqian Sun is a professor of the College of Computer Science and Technolgy at Zhejiang University which is located in Hangzhou, Zhejiang Province of China. He is now the Director of Modern Industrial Design Institute, Zhejiang University. Since 1999 his works are concentrated in the Computer-aided Industrial Design and Conceptual Design, Applied Ergonomics and Design, Virtual Human and New Medium Design etc.