

Reconhecimento de emoção de música na web com base na programação de expressão gênica mais eficaz

Kejun Zhang^a, Shouqian Sun

^aFaculdade de Ciência da Computação, Universidade de Zhejiang, Hangzhou 310027, China

informações do artigo

Disponível online em 8 de outubro de 2012

Palavras-chave:

Reconhecimento de emoção musical
Algoritmo evolutivo
Programação de expressão gênica
Máquina de vetor de suporte
Recuperação de informação musical

resumo

No estudo, apresentamos um algoritmo de maior eficácia, denominado programação de expressão gênica revisada (RGE), para construir o modelo de reconhecimento de emoção musical. Nossas principais contribuições são as seguintes: primeiramente, descrevemos os mecanismos básicos de reconhecimento de emoção musical e introduzimos a programação de expressão gênica (GEP) para lidar com a construção do modelo de reconhecimento de emoção musical. Em segundo lugar, apresentamos o RGE baseado no algoritmo evolutivo de encadeamento retroativo e usamos GEP, RGE e SVM para construir os modelos de reconhecimento de emoção musical separadamente. Os resultados mostram que os modelos obtidos por SVM, GEP e RGE são satisfatórios e bem com os valores experimentais. Por fim, relatamos a comparação desses modelos, e descobrimos que o modelo obtido pela RGE supera a precisão da classificação do modelo do GEP e leva quase 15% menos tempo de processamento do GEP e até metade do tempo de processamento do SVM, o que oferece uma nova maneira e é eficiente para resolver problemas de reconhecimento de emoção musical; ademais, como o tempo de processamento é essencial para o problema de recuperação de informação musical em larga escala, a RGE pode impulsionar o desenvolvimento da tecnologia de recuperação de informação musical.

© 2012 Elsevier BV. Todos os direitos reservados.

1. Introdução

O reconhecimento de emoções musicais, como um dos principais campos de pesquisa da recuperação de informação musical, tem sido pesquisado por muitos anos. [1], e há muitos algoritmos, frameworks e aplicativos que foram propostos na literatura [2, 3], recentemente, Yang et al. [4] propôs uma variedade hierárquica de harmonização para recuperação de mídia cruzada e apresenta uma nova estrutura para análise e recuperação de conteúdo multimídia com base em classificação semissupervisionada e feedback de relevância [5]. A obtenção de um modelo de boa qualidade para o reconhecimento da emoção musical depende de muitos fatores, como a seleção de métodos estatísticos e extração de características, para o posterior, recentemente, Ma et al. [6] propôs um novo método de seleção de recursos e o aplicou à anotação automática de imagens. No estudo, nos preocuparemos principalmente com os métodos estatísticos, embora a máquina de vetores de suporte (SVM) tenha sido escolhida como um dos melhores métodos estatísticos. [7, 8], pode ser demorado. Portanto, permanece a necessidade de um método eficiente para melhorar o desempenho do reconhecimento de emoções musicais.

Estudos mostram que as formas artificiais de inteligência são eficientes para encontrar um modelo aceitável para reconhecimento de emoção musical [9–11]. Recentemente, a programação de expressão gênica (GEP) [11], como o desenvolvimento natural de algoritmos genéticos [12, 13] e programação genética [14], foi

usado para lidar com o problema de recuperação de informação musical, especialmente, Yang et al. [15] usou GEP para resolver o problema de reconhecimento de emoção musical (para cliques musicais MIDI), e a literatura sugeriu que GEP pode melhorar a eficiência do modelo para reconhecimento de emoção musical.

No entanto, ao longo dos anos, muitas críticas ao GEP foram as seguintes:

1. Falta de método de seleção eficaz Amostragem da roda de roleta com elitismo [16] tem sido usado no GEP por muitos anos, que seja do conhecimento do autor, pouca atenção tem sido dedicada ao método de seleção do GEP [11, 15]

2. Tempo de processamento

Até recentemente, há pouca informação disponível na literatura sobre como diminuir o tempo de processamento do GEP [11, 15]

No estudo, abordamos todas essas questões e apresentamos um algoritmo mais eficaz, denominado programação de expressão gênica revisada (RGE), para construir o modelo de reconhecimento de emoção musical. A partir dos resultados experimentais, descobrimos que o modelo obtido pelo RGE supera a precisão da classificação do modelo pelo GEP e leva menos processamento. Além disso, um suporte de método de exploração difuso para "pesquisa de vetores de emoção" também será fornecido.

O artigo está organizado da seguinte forma: Seção 2, damos um briefing introdução aos métodos utilizados no estudo: GEP, RGE e SVM. Após uma introdução detalhada dos experimentos em Seção 3,

^a Autor correspondente.

Endereço de email: channy@zju.edu.cn (K. Zhang).

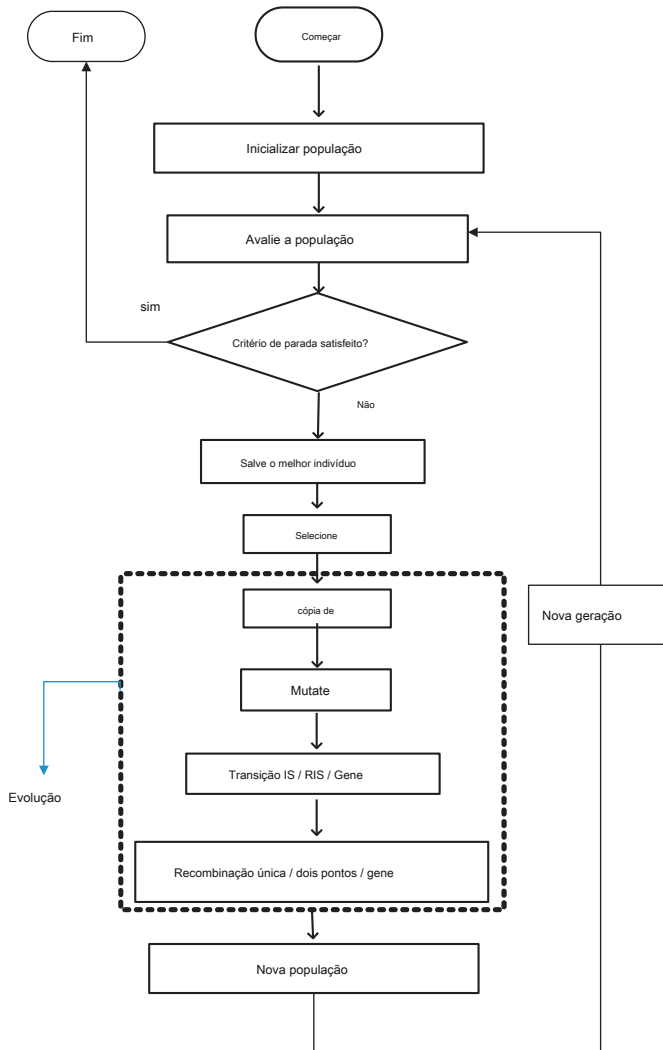


Figura 1. Programação de expressão gênica.

apresentamos os resultados experimentais e discussões em [Seção 4](#). Finalmente, resumimos as principais conclusões do estudo em [Seção 5](#).

2. Métodos

2.1. Programação de expressão gênica

GEP usa o mesmo tipo de representação de diagrama de programação genética, mas as entidades evoluídas pela árvore de expressão são a expressão de um genoma e os indivíduos são frequentemente copiados para a próxima geração com base em sua aptidão, conforme determinado pela amostragem da roda da roleta com elitismo, que garante a clonagem do melhor indivíduo para a próxima geração [11]. [Figura 1](#) mostra o fluxograma do GEP.

Como você pode ver de [Figura 1](#), GEP começa a partir de uma população inicial com muitos genes (indivíduos), após gerações de evolução, o melhor indivíduo será selecionado e seu processo de decodificação poderá ser expresso por [Fig. 2](#) [11]. No [Figura 2](#) sin representa a função sin () e terminal $x_1 - x_{10}$ representam as variáveis; os alelos representam a posição nos genes (indivíduos), como você pode ver em [Figura 2](#), de acordo com a gramática GEP [11], o indivíduo será expresso como Árvore de Expressão e a Árvore de Expressão pode ser facilmente decodificada como uma equação matemática. A descrição detalhada do GEP pode ser encontrada em [11].

Aqui, tomamos um exemplo para explicar como funciona o GEP [11]:

Etapa 1: INICIALIZAÇÃO

Definimos os indivíduos com um comprimento de 17 e o conjunto de funções:

$F \propto \{ ' + ' , ' - ' , ' \times ' , ' / ' , ' \sin ' , ' \cos ' , ' \sqrt{} ' \}$, aqui, (S representa a função sin (), C representa a função cos (), Q representa a função sqrt (). Em seguida, codificamos todas as variáveis (aqui, simplesmente, usamos apenas o 7 recursos com variáveis $x_1 - x_7$) e funções em um

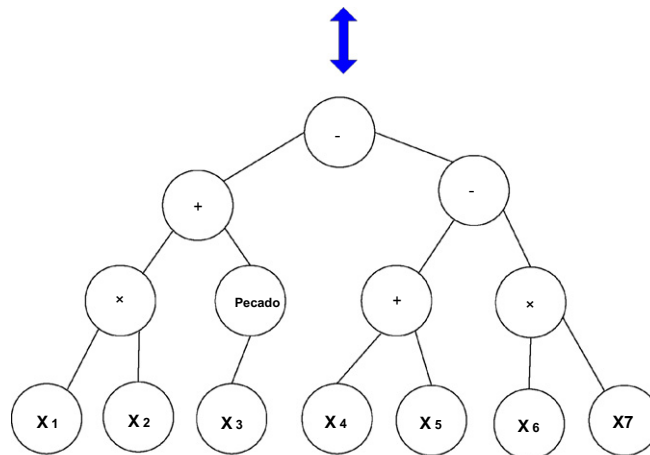
Alelo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	C	+
Valor								-	+										

Fig. 3. Indivíduo com comprimento de 17.

Equação

$$x_1 \times x_2 + \text{pecado} (x_3) - ((x_4 + x_5) - x_6 \times x_7)$$

Árvore de Expressão



Gene / Individual

Alelo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Valor	-	+															

Figura 2. Processo de decodificação de indivíduo com um gene.

9: fim se
 10: fim para
 11: Retorna População [gen] [melhor]

Algoritmo 2. Revisão da programação da expressão gênica

O RGE baseia-se na mudança da ordem de várias operações em um GEP, que oferece uma combinação de convergência rápida, maior eficiência em termos de avaliações de aptidão, equivalência estatística completa a um GEP padrão e ampla aplicabilidade [17], no entanto, requer a memorização de escolhas e indivíduos ao longo de várias gerações. De acordo com o trabalho de [17], em nossos experimentos, obtemos os seguintes resultados:

1. Para a complexidade do espaço

No pior caso possível (onde todos os programas são construídos e avaliados, portanto, o número de indivíduos efetivamente criados e avaliados durante a execução do RGE é igual ao número de indivíduos durante a execução do GEP), se o tamanho máximo dos indivíduos em cada geração ao longo da execução é de 100 bytes, só precisamos O de espaço de memória, O igual a:

$$O \propto G_{\max} \times 1 \times P \quad 100 \quad \delta 4^\circ$$

Onde G_{\max} é a geração máxima, P é o tamanho da população.

Por exemplo, se executarmos uma população de $P \propto 100$ indivíduos com geração máxima de 600 por 100 vezes, sem otimização de memória, a RGE requer apenas cerca de 101 100 600 100 bytes, 600 MB de memória para funcionar, o que está prontamente disponível na maioria dos computadores pessoais modernos.

2. Para complexidade de tempo

Para GEP, temos

$$T_{\text{GEP}} \propto M \quad \delta G \times 1 \times P \quad S_{\text{GEP}}^{\text{média}} \quad \delta 5^\circ$$

E para RGE

$$T_{\text{RGE}} \propto M \quad E_{\text{RGE}} \quad S_{\text{RGE}}^{\text{média}} \quad \delta 6^\circ$$

Obviamente, a economia de tempo proporcionada pela RGE pode ser expressa como uma Eq. (7):

$$T_{\text{Salve}} \propto \frac{1}{2} T_{\text{GEP}} T_{\text{RGE}} \propto M \quad h \quad \delta G \times 1 \times P \quad S_{\text{GEP}}^{\text{média}} E_{\text{RGE}} \quad S_{\text{RGE}}^{\text{média}} \quad \delta 7^\circ$$

Onde G é o número atual de geração, E_{RGE} é o número de indivíduos realmente criados e avaliados durante

o RGE funciona. M é a quantidade de conjuntos de treinamento, $S_{\text{GEP}}^{\text{média}}$ e $S_{\text{RGE}}^{\text{média}}$ é o tamanho médio dos indivíduos (criados) durante as execuções do GEP e do RGE, respectivamente.

Obviamente, E_{RGE} é menos do que ($G \times 1$) P; e considerando

inchar [17], $S_{\text{RGE}}^{\text{média}}$ é muito menor que $S_{\text{GEP}}^{\text{média}}$. Portanto, comparou para o GEP, muito tempo será economizado na RGE; além disso, para uma população inchada, a parcimônia do RGE em termos de avaliações de aptidão é combinada com sua parcimônia em termos de tamanhos de programas para produzir economias ainda mais impressionantes [16].

2.3. Máquina de vetor de suporte

SVM é um tipo de método de aprendizado de máquina e uma função de espaço de recurso induzida por kernel é usada para o mapeamento de objetos em valores alvo [7]. No SVM, um espaço de recursos induzido pelo kernel

com função $k(x, x_{\text{eu}})$ é usado para o mapeamento de objetos em valores de destino. Assim, um mapeamento de características não lineares permitirá o tratamento de problemas não lineares em um espaço linear. A previsão ou

Aproximadamente, a função imation usada por um SVM básico é uma Eq. (8):

$$f(\delta x) \propto \frac{1}{2} \text{eu} \quad \text{uma eu } K(\delta x, x_{\text{eu}}) \propto b \quad \delta 8^\circ$$

Onde eu é algum valor real, x_{eu} é um vetor de característica correspondente a uma objeto de treinamento, e $k(x_{\text{eu}}, x)$ é uma função do kernel. Os componentes do vetor eu e a constante b representam a hipótese e são otimizados durante o treinamento. $k(x_{\text{eu}}, x)$ é uma função kernel, cujo valor é igual ao produto interno de dois vetores x e x_{eu} no espaço de recurso $j(x)$ e $j(x_{\text{eu}})$. Neste estudo, usamos uma função de base radial gaussiana como a função kernel.

$$K(\delta u, v) \propto \exp \quad g \propto \frac{1}{2} \text{uv} \quad \delta 9^\circ$$

Onde g é constante, você e v são duas variáveis independentes, a generalização de SVM pode ser controlada ajustando a forma de gaussiana com g. Detalhes sobre SVM, consulte [7].

3. Experiências

3.1. Base de dados

Um total de 726 partes rítmicas principais (clipes musicais com 30 s) de canções populares relevantes no formato MP3 foi selecionado no estudo. De acordo com [19], usamos MARSYAS [20] para recuperar 30 recursos de cada clipe de música, que inclui 19 recursos de textura tímbrica (meios e variações de centróide espectral, roll-off, fl uxo, zerocrossings sobre a janela de textura (8), baixa energia (1) e meios e variações de os primeiros cinco coeficientes MFCC sobre a janela de textura (excluindo o coeficiente correspondente ao componente DC)) 6 de recursos de conteúdo rítmico, 5 de recursos de conteúdo de pitch [19]. Conforme relatado em [21], formação complementar em escolas de música

ou não é desprezível para determinar a habilidade musical (experiência emocional) de seres humanos; portanto, recrutamos 80 alunos para fazer anotações nos clipes musicais e não classificamos os alunos com o fundo. Para torná-lo mais objetivo, os clipes musicais que usamos são desconhecidos dos alunos. De acordo com o modelo de Hevner [22], realizamos um estudo sobre a seleção de adjetivos para o espaço da emoção musical

[23, 24], e 8 adjetivos selecionados para rotular a emoção musical: Vigoroso, Digno, Triste, Sonho, Calmante, Gracioso, Alegre e Emocionante. Em seguida, os alunos precisam anotar cada clipe de música com esses 8 rótulos de emoção, definir um valor de 0 a 1 para representar a semelhança para cada emoção (o valor próximo a 1 significa muito semelhante), por exemplo, uma música pode ser anotada como:

$$\text{Música Emoções} \propto \frac{1}{2} 0: 3,0: 2,0: 2,0: 1,0: 3,0: 2,0: 6, 0: 9 \quad \delta 10^\circ$$

onde, os valores representam a semelhança com Vigoroso, Digno, Triste, Sonho, Calmante, Gracioso, Alegre e Emocionante, respectivamente. Por exemplo, o valor de "Triste" é 0,2, significa "não está triste"; o valor de "Excitante" é 0,9, significa "muito emocionante". No estudo, a emoção com valor máximo (negrito) será definida como a emoção principal da música, portanto, a emoção principal da música de exemplo é "emocionante".

Após a anotação, a maioria dos alunos anotou todos os clipes musicais, se um clipe musical foi anotado e a emoção principal ficou com menos de 70% da anotação total, consideramos o clipe musical como difícil de distinguir e o excluímos do banco de dados. Em suma, obtemos 46.648 anotações: cada uma das 686 músicas é anotada por 68 participantes. Escolhemos o valor médio como o valor final para construir todo o banco de dados. Fig. 5 mostra o procedimento de construção de banco de dados para reconhecimento de emoção musical.

3.2. Avaliação

Para avaliar o sistema, duas equações clássicas são usadas: Erro quadrático médio (MSE) do programa (individual) eu e correlação

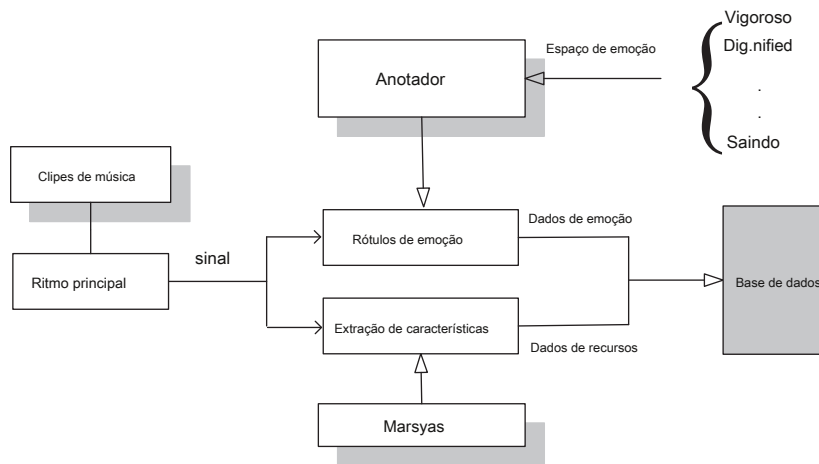


Fig. 5. Procedimento de construção de banco de dados para reconhecimento de emoção musical.

coeficiente de correlação (CC).

$$MSE_{eu} = \frac{1}{n} \sum_{j=1}^n (F(eu_j) - Q_j)^2$$

Onde $F(eu_j)$ é o valor previsto pelo programa individual eu para o estoj de aptidão j (fora de n estojs de aptidão) e Q_j é o valor alvo para o estoj de fitness j .

$$CC = \frac{Cov(F, Q)}{\sigma_F \sigma_Q}$$

Onde $Cov(F, Q)$ é a covariância das saídas alvo e modelo, σ_F e σ_Q são os desvios padrão correspondentes.

3.3. Modelo para reconhecimento de emoção musical

O reconhecimento da emoção musical é um processo de modelagem entre o espaço das características musicais e o espaço emocional (ver Fig. 6). No Fig. 6, X significa que a música apresenta espaço; Y significa espaço de emoção. $F(X)$ é o modelo esperado.

Informações detalhadas sobre modelagem para reconhecimento de emoção musical por SVM (como encontrar o melhor $F(X)$), por favor, veja [7].

Fig. 7 mostra o modelo de reconhecimento de emoção musical baseado no GEP e RGEF:

4. Resultados e discussões

Selecionamos 600 clipes de música como o conjunto de treinamento e os 86 clipes de música restantes como o conjunto de teste, e vamos nos preocupar principalmente com o reconhecimento da emoção principal. Portanto, três etapas serão realizadas para alcançar os resultados da comparação:

Etapas 1: treinar o classificador por 100 vezes separadamente com base no conjunto de treinamento e comparar o tempo de processamento relevante no GEP, RGEF e SVM;

Passo 2: Conduza os experimentos várias vezes para encontrar o melhor modelo;

Etapas 3: teste o melhor modelo no conjunto de teste.

Os detalhes serão descritos nas seções a seguir.

4.1. GEP e RGEF

No estudo, todos os parâmetros e os conceitos do GEP e RGEF são utilizados com referência à literatura. [11]. Em primeiro lugar, na etapa de treinamento, dividimos o conjunto de treinamento em 10 partes e fazemos

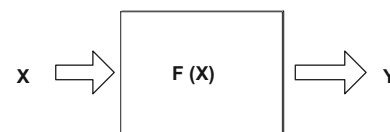


Fig. 6. Modelo para reconhecimento de emoção musical.

Teste de validação cruzada de 10 vezes. Os experimentos mostram que GEP e RGEF obtêm MSE abaixo de 0,020. Então, tomando MSE $\frac{1}{4}$ 0,020 como critério de parada, conduzimos o experimento para todos os conjuntos de dados de treinamento por GEP e RGEF por 100 vezes (na mesma condição com os mesmos parâmetros (ver tabela 1), os outros parâmetros são iguais, consulte [11]), cada execução parava apenas se o critério de parada fosse satisfeito ou além de 10 min do limite máximo de tempo de execução.

Em segundo lugar, conduzimos os experimentos com geração máxima $G \frac{1}{4}$ 600 (sem critério de parada) em todo o treinamento definido várias vezes pelo GEP e RGEF, e salvar os melhores modelos com um mínimo MSE de GEP e RGEF respectivamente (ver mesa 2). Por fim, teste o melhor modelo no conjunto de testes separadamente e obtenha os resultados de comparação em Tabela 3.

4.2. SVM

No estudo, o método foi implementado com base em WEKA e LibSVM [25, 26]. Em primeiro lugar, na etapa de treinamento, precisamos encontrar a melhor g , c , σ . De acordo com a capacidade de generalização do modelo com base na validação cruzada Leave One Out para o conjunto de treinamento, obtemos os melhores parâmetros de g , c e σ são 0,006,

0,015 e 68 respectivamente. Com esses parâmetros, tomando MSE $\frac{1}{4}$ 0,020 como critério de parada, conduzimos os experimentos em todo o conjunto de treinamento por 100 vezes (o mesmo que GEP e RGEF).

Em segundo lugar, realizamos os experimentos com limites de tempo de 10min (sem critério de parada) em todo o treinamento definido várias vezes pelo SVM, e salvamos o melhor modelo com um mínimo MSE (Veja mesa 2). Finalmente, teste o melhor modelo no conjunto de teste e obtenha os resultados de comparação em Tabela 3.

4.3. Discussões

No estudo, todos os experimentos são conduzidos em Win7 com Intel core 2 Duo 2,53 G e 4 G de memória. Como você pode ver de mesa 2, RGEF supera a precisão de classificação do GEP e leva menos tempo de processamento (84,62% do GEP e quase metade do SVM).

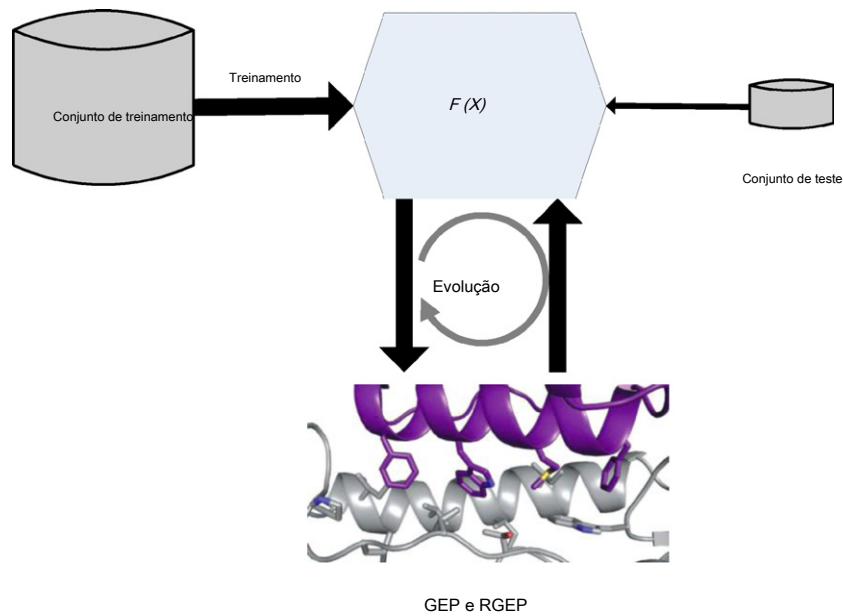


Fig. 7. Modelo de reconhecimento de emoção musical baseado em GEP e RGEP.

tabela 1
Parâmetros usados no GEP e RGEP.

Parâmetros	Valor
Tamanho da população	100
Conjunto de funções terminal	$F \frac{1}{4} [p, \dots, C, S, C, Q]$
Números de genes, tamanho da cabeça do gene Taxa de mutação	5, 15 0,044
Combinação de um / dois pontos	0,2
Recombinação gênica	0,05
IS / RIS / Taxa de transposição do gene	0,05

mesa 2
Resultado da comparação do reconhecimento da emoção principal da música (Conjunto de treinamento).

Algoritmo	CC	MSE	100 tempo (s) de execução
SVM	0,782	0,0072	31.782
GEP	0,6425	0,0193	15.892
RGEP	0,7602	0,0157	13.448

Tabela 3
Resultados do reconhecimento da emoção principal da música (conjunto de teste).

Algoritmo	Taxa de sucesso (%)			
	$F \frac{1}{4} 0,005$	$F \frac{1}{4} 0,010$	$F \frac{1}{4} 0,015$	$F \frac{1}{4} 0,020$
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

Nós podemos ver de Tabela 3, os melhores resultados de reconhecimento são 87,21%, 84,88% e 87,21% em SVM, GEP e RGEP, respectivamente. E o melhor valor para F pode ser entre 0,010 e 0,015 para eles.

Aqui, um limite d é usado, d pode ser definido como

$$d \frac{1}{4} 9 F Q 9$$

$$\delta 13^{\circ}$$

F é o valor previsto e Q é o valor experimental relevante. Então, definimos um limite F, E se d o F, nós o consideramos um teste de sucesso.

Para ser mencionado, em trabalho recente, adicionamos mais conjuntos de funções ao RGEP, a precisão da classificação é extremamente melhorada, porém, o tempo de processamento também aumentou. Portanto, as pesquisas futuras devem prestar mais atenção ao melhor achado dos conjuntos de funções em termos de tempo de processamento.

Além disso, gosto [24], já que o sistema RGEP leva a música cliques como um vetor de emoção; portanto, uma música com 8 rótulos de emoção: Vigorous, Dignified, Sad, Dreaming, Calothing, Graceful, Joyous e Exciting, poderia ser calculada como a Eq. (14):

$$\text{Música Emoções } \frac{1}{4} F \delta x^{\circ} 1, F \delta x^{\circ} 2, F \delta x^{\circ} 3, F \delta x^{\circ} 4, F \delta x^{\circ} 5, F \delta x^{\circ} 6, F \delta x^{\circ} 7, F \delta x^{\circ} 8 \quad \delta 14^{\circ}$$

Onde $F(x)1 - F(x)8$ representam o valor de Vigoroso, Digno, Triste, Sonho, Calmante, Gracioso, Alegre e Emocionante.

Conseqüentemente, podemos obter alguma exploração difusa usando este sistema. Por exemplo, para uma música com vetor de emoção:

$$\text{Música Emoções } \frac{1}{4} 0: 3,0: 2,0: 2,0: 1,0: 3,0: 2,0: 6,0: 9 \quad \delta 15^{\circ}$$

Podemos explorar essa música pelas palavras-chave " muito emocionante e alegre " (em negrito) na web. É uma forma muito boa e eficaz de pesquisar músicas.

5. Conclusão

No estudo, apresentamos uma programação de expressão gênica revisada com base no algoritmo evolutivo de encadeamento reverso para construir o modelo de reconhecimento de emoção musical. A partir dos resultados experimentais, descobrimos que o modelo obtido pela RGEP supera a precisão de classificação do modelo GEP e leva quase 15% menos processamento de GEP e até metade do tempo de processamento de SVM, o que oferece uma nova maneira e eficiente para resolver problemas de reconhecimento de emoção musical; E como o tempo de processamento é essencial para o problema de recuperação de informação musical em larga escala (MIR), a RGEP pode levar ao desenvolvimento da tecnologia MIR da web. Além disso, um suporte de método de exploração difuso para " pesquisa de vetores de emoção " também é fornecido.

Pesquisas adicionais devem incluir uma avaliação mais detalhada do desempenho do sistema para conjuntos de dados em grande escala, encontrar um método de seleção de recurso apropriado, pesquisa sobre reconhecimento de rótulo múltiplo e também o parâmetro F pode ser aplicado na função de aptidão mais tarde.

Agradecimentos

Este estudo é parcialmente financiado pela Fundação Nacional de Ciências Naturais da China (61070075, 61004116, 61003147).

Referências

- [1] X. Hu, JS Downie, C. Laurier, M. Bay, A. Ehmann, The 2007 MIREX audio tarefa de classificação de humor: lições aprendidas, em: Anais do 9º Simpósio Internacional sobre Recuperação de Informação Musical (ISMIR), setembro de 2008, Filadélfia, PA.
- [2] D. Ververidis, C. Kotropoulos, Fast sequential floating forward selection aplicada às características da fala emocional estimadas nas coletas de dados DES e SUSAS, em: Proceedings of the XIV European Signal Processing Conference, 2006.
- [3] MM Ruxanda, BY Chua, A. Nanopoulos, CS Jensen, Música baseada em emoções recuperação em um espaço de recurso de áudio bem reduzido, em: Proceedings of the 34th International Conference on Acoustics, Speech, and Signal Processing, 19-24 de abril de 2009, Taipei, Taiwan, 4p, no prelo.
- [4] Yang yi, Yue-Ting Zhuang, Fei Wu, Yun-He Pan, Harmonização hierárquica manifolds para compreensão da semântica de documentos multimídia e recuperação de mídia cruzada, IEEE Trans. Multimedia 10 (3) (2008) 437–446.
- [5] Yang Yi, Nie Feiping, Xu Dong, Luo Jiebo, Zhuang Yueting, Pan Yunhe, A framework de recuperação de multimídia baseado em classificação semissupervisionada e feedback de relevância, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 723–742. [6] Zhigang Ma, Feiping Nie, Yi Yang, Jasper Uijlings, Nicu Sebe, imagem da web anotação por meio de seleção de recursos colaborados de dispersão de subespaço, IEEE Trans. Multimedia, <http://dx.doi.org/10.1109/TMM.2012.2187179>, no prelo.
- [7] C. Cortes, V. Vapnik, Redes de vetores de suporte, Mach. Aprender. 20 (1995) 273–297.
- [8] Jianhua xu, uma máquina de vetor de suporte estendido um versus resto para multi classificação de rótulo, Neurocomputing 74 (17) (2011) 3114–3124.
- [9] S.-B. Cho, imagem emocional e recuperação de informação musical com interação algoritmo genético, Proc. IEEE 92 (4) (2004) 702–711.
- [10] MH Sedaaghi, D. Ververidis, C. Kotropoulos, Melhorando a emoção da fala reconhecimento usando algoritmos genéticos adaptativos, EURASIP (2007) 2209–2213. [11] C. Ferreira, Programação de expressão gênica: um novo algoritmo adaptativo para resolução de problemas, Complex Syst. 13 (2003) 87–129.
- [12] JH Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.
- [13] DE Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley, Boston, Massachusetts, Estados Unidos, 1989.
- [14] JR Koza, Genetic Programming: On the Programming of Computers, por Means of Natural Evolution, MIT Press, Massachusetts, 1992.
- [15] C. Yang, S. Sun, K. Zhang, T. Liu, Estudo sobre o modelo de cognição da emoção musical baseado na aplicação da programação de expressão gênica aprimorada, em: Proceedings of the Second Workshop on Digital Media e sua aplicação em Museum & Heritage, 2007, pp. 344–351.
- [16] M. Mitchell, Uma introdução aos algoritmos genéticos, Complex Adaptive Systems, MIT Press, Cambridge, Massachusetts, Estados Unidos, 1996.
- [17] R. Poli, WB Landon, algoritmos evolutivos de encadeamento reverso, Artif. Intell. 170 (2006) 953–982.
- [18] T. Back, DB Fogel, T. Michalewicz (Eds.), Evolutionary computation Q: Basic Algorithms and Operators, Institute of Physics Publishing, 2000.
- [19] G. Tzanetakis, P. Cook, Classificação do gênero musical de sinais de áudio, 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, Reino Unido, 2000. [21] E. Bigand, B. Poulin-Charronnat, Are we experiencing aural? Uma análise de as capacidades musicais que não dependem do treinamento musical formal, Cognition 100 (2006) 100–130.
- [22] K. Hevner, Estudos experimentais dos elementos de expressão na música, Am. J. Psychol. 48 (1936) 246–268.
- [23] Tao Liu, Pesquisa em Modelo de Reconhecimento de Emoção Musical e Interativo Tecnologia [D], Universidade de Zhejiang, Hangzhou, China, 2006.
- [24] Bin Zhu, Kejun Zhang, sistema de reconhecimento de emoção musical baseado em GA-BP, em: Proceedings of the 2010 International Conference on Computer Design and Applications (ICDDA), vol. 2, 2010, V2-409 – V2-412.
- [25] I. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2ª edição, Morgan Kaufmann Publishers Inc., San Francisco, 2005. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/> S.
- [26] C.-C. Chang, C. Lin, LIBSVM: a library for support vector machines, 2001.



Kejun Zhang recebeu Ph.D. em Ciência da Computação pela Universidade de Zhejiang, China. Atualmente é pós-doutorado na Faculdade de Ciência da Computação da Universidade de Zhejiang, na China. Seus interesses de pesquisa estão em recuperação de informação musical, bioinformática, computação evolutiva e aprendizado de máquina. Ele é membro do IEEE. Ele publicou muitos artigos de pesquisa em várias revistas e anais de conferências de renome.



Shouqian Sun é professor da Faculdade de Ciência da Computação e Tecnologia da Universidade de Zhejiang, localizada em Hangzhou, província de Zhejiang na China. Ele agora é o Diretor do Instituto de Design Industrial Moderno da Universidade de Zhejiang. Desde 1999 os seus trabalhos concentram-se em Desenho Industrial e Desenho Conceptual Assistido por Computador, Ergonomia e Desenho Aplicado, Humano Virtual e Design de Novos Médios, etc.