

Transcrição Automática de Música

Ricardo Rosa e Miguel Eliseu

Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria

Departamento de Engenharia Informática

A transcrição automática de música é um problema computacionalmente difícil que a esta altura continua ainda sem solução. As diversas abordagens têm sido sobretudo centradas no conhecimento teórico que existe do processamento de sinal. Apesar de o conhecimento sobre esta matéria ter evoluído muito nos últimos anos, assim como os meios tecnológicos para o seu estudo, o sucesso tem sido relativo[2] (poucas notas, centrada no mesmo instrumento). Perante as dificuldades encontradas na transcrição a partir do sinal áudio, surgiu a ideia de procurar a solução partindo da transcrição para o sinal áudio. Mais concretamente, achar a transcrição que melhor represente o sinal áudio em análise, tirando partido do sucesso que os algoritmos genéticos têm tido na resolução de problemas complexos e de espaços de pesquisa muito grandes, intratáveis pelos tradicionais métodos de pesquisa. O método apresentado é válido, tal como é demonstrado nos resultados finais.

Introdução

A transcrição automática de música é o processo de transcrever música para uma pauta, segundo uma determinada notação (símbolos que representam notas musicais, tempos, etc.). Neste contexto entende-se por transcrição automática de música quando um programa ou um pedaço de software escreve as partituras de cada instrumento.

O problema de transcrição automática de música, quase nunca foi visto como um problema de pesquisa onde o objectivo passa por criar uma sequência de notas que reproduza o sinal áudio original. O facto do espaço de pesquisa ser vasto e exponencial à medida que adicionamos notas e instrumentos, tem levado a que este paradigma ainda não tenha sido muito explorado no contexto da transcrição de música. Os algoritmos genéticos, por definição, são algoritmos direccionados para problemas em que o espaço de pesquisa é muito grande.

Algoritmos Genéticos

Os algoritmos genéticos, baseiam-se em processos biológicos da genética e na evolução através da selecção natural. Uma perspectiva sustentada na analogia da forma como as espécies evoluem. De acordo com a teoria proposta por *Charles Darwin* [3] e complementada por Mendel, as espécies evoluem graças a um processo de selecção natural, processo este que promove a sobrevivência dos indivíduos mais adaptados.

Durante a reprodução, o material genético dos indivíduos está sujeito a modificações determinadas por operadores genéticos, como sejam a recombinação e a mutação (detalhados mais à frente). Estes operadores

provocam diferenças genéticas que se podem tornar vantajosas em determinadas condições, tornando os indivíduos mais aptos, conferindo-lhes uma maior probabilidade de sobreviver.

Através de um processo iterativo ao longo de várias **gerações**, o algoritmo faz evoluir uma **população** de soluções candidatas para um dado problema. Cada **indivíduo** da população contém material genético, designado de **genoma**. O genoma é constituído por um ou mais **cromossomas**. A unidade hereditária de um cromossoma é o **gene** e codifica determinada característica do indivíduo. Os valores que cada gene pode tomar, designa-se por **alelos**. Ao conjunto particular de genes contidos no genoma designamos **genótipo**. A representação visível das características do indivíduo, codificada no genótipo, designa-se por **fenótipo**. O **mérito, qualidade ou fitness** de um indivíduo traduz a capacidade que ele tem de sobreviver e produzir descendência, reflectindo a aproximação do indivíduo à solução do problema. A qualidade é medida por uma função de avaliação. O mecanismo de **recombinação** consiste na troca de material genético entre dois indivíduos. Os novos indivíduos resultantes do processo de recombinação podem ser sujeitos a **mutações** que provocam alterações dos valores de um ou mais genes.

Algoritmo genético no problema da Transcrição Automática de Música

Representação genética

No modelo adoptado um gene representa uma nota (o silêncio foi tratado como mais uma nota). Um gene é uma estrutura de dados constituída por um *pitch*, uma duração e a máscara (o papel da máscara é detalhado na descrição da função de avaliação).

O *pitch* pode tomar valores de 0 a 129 (em que o silêncio é o 0) e os restantes valores representam as 128 notas do formato MIDI.

A duração toma o valor das durações. Cada valor representa uma duração standard.

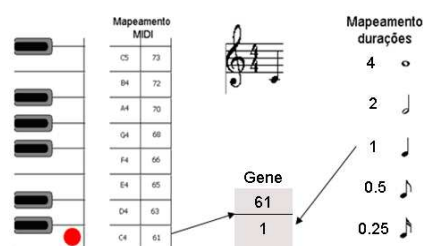


Figura 1. Um gene representa uma nota

A sequência dos genes no cromossoma é sequência das notas na pauta.



Figura 2. Frase Musical

Um cromossoma é um conjunto de genes, que representam uma frase musical ao qual se adiciona o atributo que define um instrumento. O genótipo do indivíduo é constituído por um conjunto (parametrizável, mas igual para todos os indivíduos) de cromossomas que representam uma pauta musical.

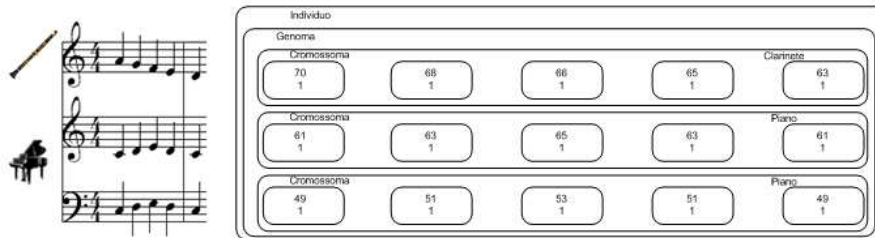


Figura 3. Pauta Musical

Uma população é um conjunto de indivíduos que são candidatos à solução correcta.

Geração da população inicial

A população inicial é gerada de forma aleatória. É definido o tamanho máximo, em termos de duração, que um indivíduo pode ter, e ainda quantos cromossomas, formarão o genótipo de cada indivíduo.

Método de Selecção

A selecção por torneio foi o método utilizado que vai seleccionando dois indivíduos da população aleatoriamente, com reposição, para torneio e devolve para recombinação 50% do tamanho da população.

Recombinação

Na implementação os indivíduos podem ter tamanhos diferentes. Um ponto no tempo é escolhido aleatoriamente, sendo o tempo máximo, o mínimo da duração dos indivíduos e do áudio original. O ponto de corte é ajustado cromossoma a cromossoma de modo as que notas passem inteiras de um indivíduo para o outro. Neste trabalho 90% da população é recombinada entre si.

Mutação

A mutação opera em todos os genes de um cromossoma, ou seja, todos os genes têm uma determinada probabilidade de serem mutados e actua sobre todos os indivíduos da população. Os operadores estão divididos em duas categorias:

- Mutações sobre o pitch, seja subir ou descer um pitch, uma oitava, copiar à direita ou à esquerda, colocar silêncio, alterar aleatoriamente, trocar de genes ou instrumentos entre cromossomas
- Mutações sobre a duração, seja aumentar ou diminuir, agregar à direita ou à esquerda, dividir uma nota

Função de Avaliação

Da pesquisa efectuada sobre a análise de sinal surgiu a ideia de utilizar o coeficiente de correlação de dois vectores. A correlação é um método de comparação que permite determinar a semelhança (ou diferença) entre dois sinais, baseado na média do seu produto temporal.

Este coeficiente varia entre os valores -1 e 1. O valor 0 (zero) significa que não há relação linear, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita mas inversa, ou seja quando uma das variáveis aumenta a outra diminui. Quanto mais próximo estiver de 1 ou -1, mais forte é a associação linear entre as duas variáveis.

O coeficiente de correlação é obtido de forma iterativa sobre a FFT calculada sobre pequenos pedaços de tempo.

O pseudo código seguinte mostra a função de avaliação, onde se evidencia a aplicação deste método:

```
Erro = N°PedaçosFFTOOriginal2
Criar vector mascara com tamanho igual a N°PedaçosFFTOOriginal
TotalDif=0
I=0
Por cada PedaçoFFTOOriginal
  Coef=-1;
  PedaçoTeste = NovoPedaçoFFTTeste
  Caso ainda existe algo para ler do Teste
    Coef=CalculaCoeficienteCorrelacao(PedaçoFFTOOriginal,PedaçoTeste)
  Fim de Caso
  Caso o coef>0.9975
    Erro=Erro- N°PedaçosFFTOOriginal
    Mascara na posicao I = 1
  Fim de Caso
  TotalDif=TotalDif+1-Coef
  I=I+1
Fim Por cada
Actualiza a máscara do indivíduo de teste
Devolve Erro+TotalDif/ N°PedaçosFFTOOriginal
```

O valor da função de avaliação tenderá para 0, em 0 é o indivíduo perfeito.

A máscara tem a responsabilidade de guardar informação sobre os pedaços que estão correctos no indivíduo de teste.

Pedaços Certos

```
Instrument: 0
Pitch: -1,32,-1,99,
Duration: 0.75,0.25,0.5,0,
GeneChainMask: [1,1,1,1,1]0,0,0,0,0,0,0,1,1,1,1,1]0,0,0,0,0,0,0,0,0,0,1,] [27]
```

Figura 4. Máscara de um indivíduo

Cada pedaço equivale a um espaço temporal. Neste caso, em que a FFT é feita recorrendo pedaços de 2048 amostras, o espaço temporal é de 0.09 segundos, o que equivale a ter sensivelmente 5 pedaços para uma nota com duração igual 0,5.

População submetida à Recombinação

Tal como referido no método de selecção, 50% da população tem origem no torneio. Os outros 50% são baseados na informação genética do melhor indivíduo da geração anterior. É aqui que é utilizada a máscara descrita no ponto anterior. O algoritmo procura manter nos novos indivíduos as partes que estão certas e opera mutações (são utilizados os operadores de mutação) sobre os pedaços maus.

Às notas é-lhes definido um valor da máscara baseado no número de 1s que lhe pertence. Esse valor é dado pela forma

$$X = \frac{c}{s-2}$$

em que c é o número 1s e o s é o tamanho da máscara atribuída à nota. Ao s é-lhe retirado 2 devido a que a nota pode estar no sítio certo e as notas adjacentes estarem erradas, e na fronteira a máscara terá o valor de 0.

```

Instrument: 0
Pitch: -1,32,-1,99,
Duration: 0.75,0.25,0.5,1.0,
GeneChainMask: [1,1,1,1,1,0,0,0,0,0,0,1,1,1,1,0,0,0,0,0,0,0,0,0,0,1,] [27]
Mask: [1,1,1,1,1,0,0,0,]»0.8333333[8], [0,0,0,]»0.0[3], [1,1,1,1,1,]»1.0[5], [0,0,0,0,0,0,0,0,0,0,1,]»0.11111111[11],
    
```

Máscara atribuída à nota Valor da Máscara

Figura 5. Máscara de um gene

Quando a máscara tem o valor de 1, significa que a nota e a duração estão correctas naquela posição, contudo é possível que a duração correcta seja maior. De modo a assegurar esta situação, com probabilidade baixa (a da mutação), é seleccionada aleatoriamente uma das mutações sobre a duração e executada sobre este gene. Quando a máscara tem o valor 0 ou um valor equivalente à articulação da nota (silêncio no final da nota) significa que nem a nota nem a duração estão correctas, então é seleccionada aleatoriamente uma das mutações do *pitch* e executada sobre este gene. Para os outros casos, o gene é partido e novos genes com as partes certas e partes erradas, tal como a imagem seguinte demonstra. Em que às partes com máscara a 0 é seleccionado um *pitch* aleatório.

```

Instrument: 0
Pitch: 32,32,-1,
Duration: 0.75,0.25,1.75,
GeneChainMask: [0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0,0,0,0,0,0,0,0,0,0,1,] [27]
Mask: [0,0,0,0,0,0,0,0,]»0.0[8], [0,0,0,]»0.0[3], [1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,1,]»0.42857143[16],
*****
Geração 4
Melhor do RUN: Fitness: 432.58262931856314 Size: 1 - A cromossome with 1 gene's chains

Instrument: 0
Pitch: -1,32,-1,99,
Duration: 0.75,0.25,0.5,1.0,
GeneChainMask: [1,1,1,1,1,0,0,0,0,0,0,0,1,1,1,1,0,0,0,0,0,0,0,0,0,1,] [27]
Mask: [1,1,1,1,1,0,0,0,]»0.8333333[8], [0,0,0,]»0.0[3], [1,1,1,1,1,]»1.0[5], [0,0,0,0,0,0,0,0,0,0,1,]»0.11111111[11],
    
```

Figura 6. Divisão do gene com base na máscara

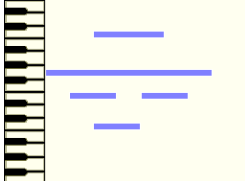
Critério de paragem

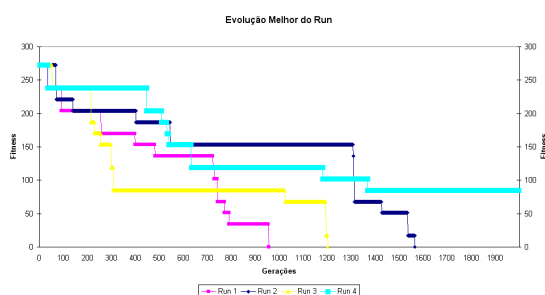
O critério de paragem é quando o melhor indivíduo tem o valor de avaliação igual a zero ou a máscara toda a 1s. Apesar, da tendência do algoritmo, encontrar sempre a solução óptima é também definido outro critério de paragem baseado no número de gerações.

Testes e Análise dos Resultados

Foram executados vários testes polifónicos em registos de frases musicais distintas, quer no desenho da frase quer no “enredo” de figuras musicais utilizadas, testes que testes que procuraram validar a robustez do método e analisar a qualidade das soluções encontradas. De seguida é apresentado um dos testes realizados e a respectiva evolução das gerações ao longo de 4 runs distintos.

Tabela 1. Cenário de Teste

	Inst		Pitch/Duração	
	0	REST/0.5	C4(60)/0.5	REST/0.25
1	REST/0.25	REST/0.25	E4(64)/0.5	E4(64)/0.5
2	G4(67)/1.5			
3	REST/0.5	C5(72)/0.75	REST/0.25	



Perante este cenário o algoritmo foi capaz de chegar por três vezes à solução correcta. Os resultados obtidos revelam que a abordagem proposta é válida e permite ambicionar melhores resultados.

Referências

- [1] Costa, E., Simões, A., Inteligência Artificial – Fundamentos e Aplicações, FCA, 2004.
- [2] Klapuri, A., Automatic Music Transcription as We Know it Today in Journal of New Music Research 2004, Vol. 33, No. 3, pp.269-282.
- [3] Darwin, C., A origem das espécies, 1859.