

Usando Agentes Autônomos para Improvisar Composições Musicais em Tempo-Real

Artigo feito por: Patrick Hutchings(B) e Jon McCormack
Apresentado por: Paulo Fontana

Introdução

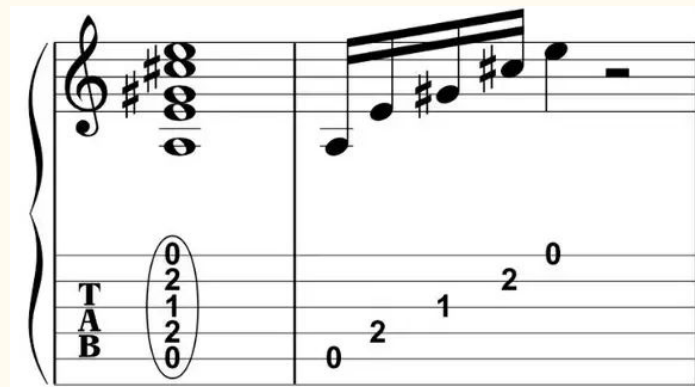


Improvisadores Virtuais

Gerar músicas originais em tempo real para performance ao vivo ou arranjos dinâmicos para Mídia como jogos apresenta muitos desafios únicos. A música deve se adaptar à mudança de humor, garantindo resultados musicalmente consistentes.

Este artigo apresenta uma abordagem para composição em tempo real de forma "vertical", usando dois agentes virtuais:

- Improvisador de **Harmonia**
- Improvisador de **Melodia**



Improvisadores Virtuais

Essa abordagem foi desenvolvida para criar um sistema de performances originais que demonstra estruturas harmônicas e consistência melódica comuns, aliados a uma adaptabilidade em tempo real.

A ideia de dividir em dois papéis diferentes (Harmonia e Melodia) vem de como essa divisão é feita em pequenos grupos de *jazz*, onde os músicos se separam em:

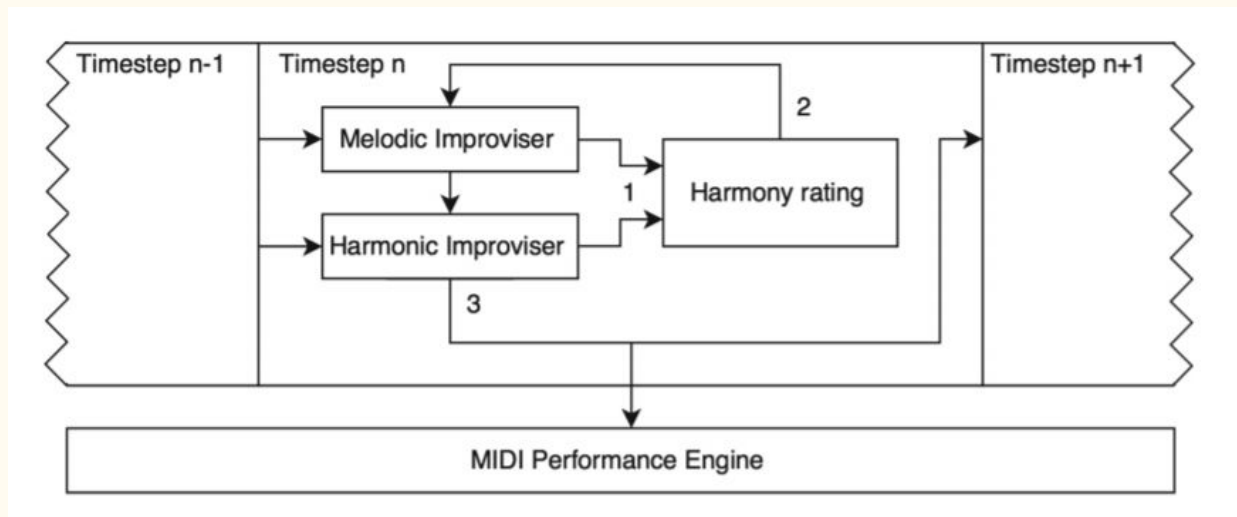
Harmonia/Rítmico (Piano, Guitarra, Baixo) e **Melodia/Solo** (Voz e Sopro).

Improvisadores Virtuais

O sistema é iniciado com melodias feitas por uma influência humana. Os agentes em seguida, exploram combinações baseadas em padrões de músicas de *jazz* conhecidas.

Por fim, os agentes auto avaliam os arranjos propostos para passar a frente à **Melodia** ou à

Harmonia



Visão geral do sistema de composição por time-step com fluxo de dados após os caminhos numerados. A música é executada e a sequência harmonizada é usada como entrada para o próximo passo de tempo.

Contexto

Harmonia e Melodia

Em músicas melódicas, melodia e harmonia são **dependentes** entre si.

Harmonia, sendo explícita ou implícita, traz contexto para as frases melódicas e tem efeitos dramáticos em tornar a melodia agradável. Em compensação, a estrutura melódica de uma composição pode sugerir um avanço harmônico já feito ou até sugerir uma nova adição.

Essa relação se torna um enorme desafio quando se trata de transpor essa ideia em algoritmos, o que dificulta tornar composições virtuais mais “humanas”. O desafio é maior ao gerar música em tempo real, pois envolve estruturas de duração variada que requer consideração além dos próximos momentos da música.

Jazz e Improvisação

Nesse caso, é importante criar um sistema se baseando em estilos musicais que envolvem a criação de melodias em tempo real através do improviso como *Blues* e *Jazz*.

No jazz, é típico que os músicos desempenhem papéis diferentes em um processo de improvisação colaborativa. Quando levamos em consideração profissionais, esses papéis não são fixos, o que torna o *Jazz* interessante é essa troca dinâmica entre quem cuida da Harmonia e quem cuida da Melodia. Mas ainda assim, em algum momento, o músico irá focar em fazer o papel da melodia ou do acompanhamento.

Jazz e Improvisação

O *Jazz* é baseado geralmente em sua estrutura harmônica. Uma sequência de acordes que irá se repetir, onde cada músico improvisa melodias que combinam com essa mesma sequência.

A sessão rítmica lida com o tempo, tom e a pegada rítmica através dos seus ouvidos e de sua intuição, podendo alterar essas partes em tempo real.

Solista reformula ou substitui melodias, mas a forma fundamental da estrutura harmônica é geralmente mantida.



Projetos Relacionados

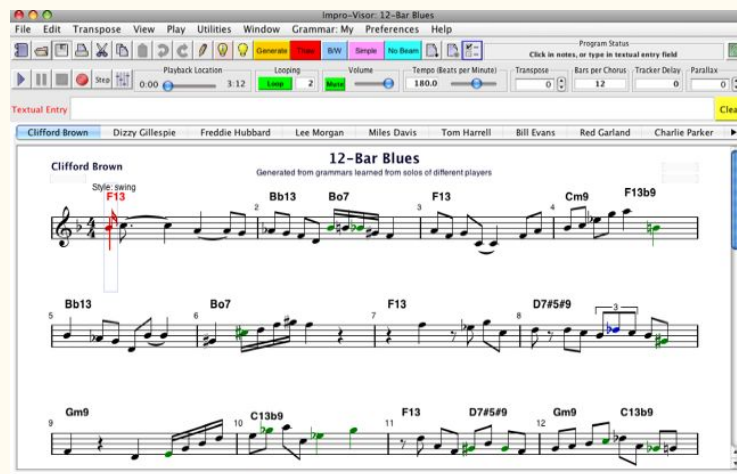
—

Improviso Computadorizado

Várias abordagens têm sido usadas para gerar música em tempo real, tanto como uma composição independente quanto para uso interativo com artistas humanos.

Impro-visor é um pacote de software que pode ser usado para gerar melodias de *jazz* usando probabilidades gramaticais. Gramáticas diferentes treinadas em solos de músicos de jazz bem conhecidos para gerar melodias em tempo real, demonstrando qualidades idiomáticas no *jazz*.

GenJammer de Biles usa um algoritmo genético para produzir frases melódicas de estilo *jazz*.

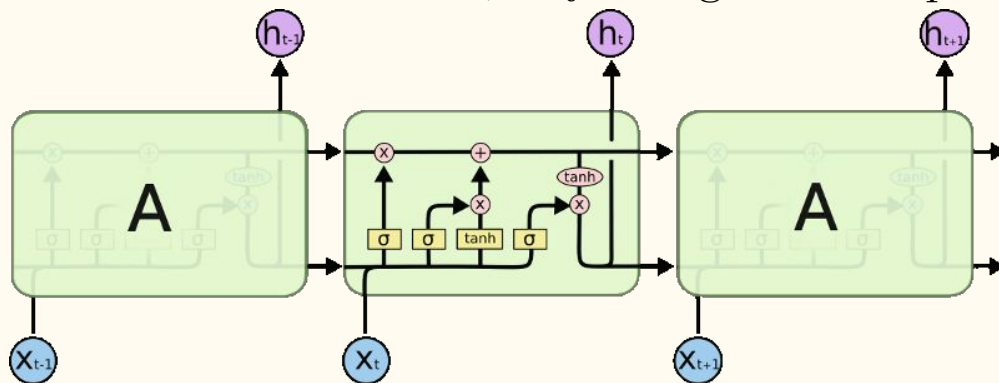


The screenshot displays the 'Impro-visor: 12-Bar Blues' software interface. The window title is 'Impro-visor: 12-Bar Blues'. The interface includes a menu bar (File, Edit, Transpose, View, Play, Utilities, Window, Grammar, My, Preferences, Help) and a toolbar with icons for file operations and playback. Below the toolbar, there are controls for 'Playback Location' (0:00 to 3:12), 'Looping' (2), 'Volume' (180.0), and 'Tempo (Beats per Minute)' (120.0). A 'Textual Entry' field is present with a 'Clear' button. The main display area shows a musical score for 'Clifford Brown 12-Bar Blues', generated from grammars learned from solos of different players. The score is in 4/4 time and features a 'Style: swing' setting. The melody is written on a treble clef staff with a key signature of one flat (Bb). The notes are color-coded: red for quarter notes, green for eighth notes, and blue for sixteenth notes. Chord symbols are placed above the staff: F13 (measures 1-2), Bb13 (measure 3), Bo7 (measures 4-5), F13 (measures 6-7), Cm9 (measure 8), F13b9 (measures 9-10), Bb13 (measure 11), Bo7 (measure 12), F13 (measures 13-14), D7#5#9 (measures 15-16), Gm9 (measures 17-18), C13b9 (measures 19-20), F13 (measures 21-22), D7#5#9 (measures 23-24), Gm9 (measures 25-26), and C13b9 (measures 27-28).

Redes Neurais LSTM (Long Short-term Memory)

Redes neurais LSTM são uma classe de redes neurais recorrentes que usam camadas de gates para simular a memória. São quatro redes neurais que compõem os gates:

- **Forget Gate** - Primeira parte, onde é decidido quais valores devem ser esquecidos (estado celular) e quais ficam.
- **Input Gate** - Cuida de atualizar os valores que vão ser usados no próximo passo.
- **Output Gate** - Decide o estado final (função “sigmoid” output + função “tanh”)



Redes Neurais LSTM (Long Short-term Memory)

O mecanismo de memória nas redes LSTM é bem adequado para aprender sequências temporais, como na música. Já sendo utilizado para criação de melodias e até na produção de novas sequências de acordes baseados em teorias de *Jazz*.

Conjuntos de dados recentemente maiores foram explorados para o treinamento de redes LSTM. O uso de unidades de processamento gráfico (GPUs) diminuiu significativamente o tempo de treinamento e execução de redes neurais profundas, executando cálculos para milhares de neurônios artificiais em paralelo e o hardware atual é capaz de treinar redes em conjuntos de dados com milhões de tokens em horas ou dias. Essa eficácia demonstrada nas redes LSTM para composição em um estilo musical específico inspirou a exploração de redes LSTM na pesquisa.

Fundamentação



Adaptar Conteúdo feito por Humanos

A capacidade de combinar as habilidades de compositores profissionais com a variabilidade dinâmica de organização controlada por algoritmos em tempo real tem aplicações diretas na classificação de mídia interativa.

Um compositor profissional pode desenvolver temas musicais baseados no conhecimento do contexto mais amplo do ouvinte, ambiente de escuta e conteúdo visual associado ou narrativa, que é difícil de simular algoritmicamente.

Adaptar Conteúdo feito por Humanos

Aumentar uma peça de música composta profissionalmente com um meio inteligente de adaptação em tempo real é desejável se a adaptação não tiver um custo significativo para a qualidade geral da peça.

O uso de manipulações de melodias comuns em melodias pré-compostas fornece um meio de adaptação com alta confiança na consistência musical em estilo e qualidade. É também um dispositivo de improvisação observado no jazz.

LSTM

Uma vantagem de usar um modelo LSTM com composições de jazz é que os padrões de jazz normalmente usam sequências de acordes repetidas de várias centenas de compassos, um comprimento adequado aos modelos LSTM que podem ser treinados e executados no hardware atual. As sequências de acordes de jazz fornecem novos desafios como um corpus de treinamento em comparação com outros gêneros de música, como folk e pop, devido à variedade de tipos de acordes e à abordagem flexível da tonalidade.

Enquanto composições inteiras podem ser criadas por amostragem de uma única rede, tais sistemas têm controles limitados em tempo real e continuam a não ter a consistência melódica e rítmica que os compositores humanos podem alcançar com relativa facilidade.

LSTM

Outras vantagens de se usar LSTM para improviso em tempo real:

- Trabalhar com mais de uma camada de rede neural para harmonia e melodia.
- Capacidade de oferecer resultados baseado em momentos (time-steps).
- Redes neurais artificiais profundas podem capturar um maior alcance estilístico dentro de uma única rede, devido a seus maiores mecanismos de tamanho e memória.
- O mecanismo de memória nas redes LSTM também aumenta o potencial de aprendizagem de relacionamentos mais distantes e dependentes do tempo nas estruturas de composição.
- Grandes resultados no processamento de linguagem, que podem ser transferidos para linguagem musical.

Improvisador Harmônico

Método

A Harmonia é dividida em estruturas de tempo e tom, nesse âmbito podemos ter mudanças de tempo e tom, o que é comum no *jazz* mas devido a sua característica de poder fazer essas mudanças em diferentes momentos, é necessário não só simplificar as opções de acordes e ritmo, mas também tratar esses acordes como algo por batida e não por composição. Mesmo sendo por batida do tempo, ainda é necessário seguir um certo padrão para evitar repetição excessiva e trocas bruscas de acordes, a cadência se torna algo a se levar em conta.

Usando como base livros regras e teorias de *jazz* se divide os acordes baseando-se nas 12 notas utilizadas na música ocidental, assim é processado os valores usando LSTM, para resultar quais entram no grupo validados, com 10% sendo aleatórios.

Método

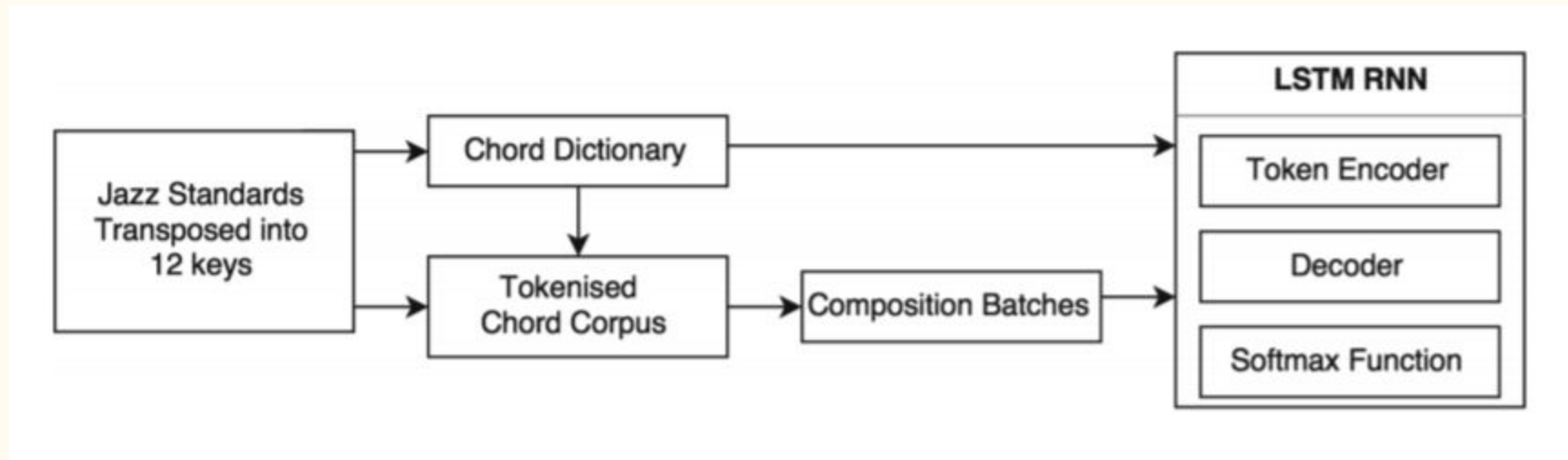


Figura: Treinando a rede neural LSTM em padrões de jazz

Resultados

Mudanças na quantidade de camadas e dimensões causaram diferentes precisões nas previsões oferecidas no LSTM.

Após testes de diferentes formas de codificar baseado na predição de confiança em 20 epochs, foi chegado a conclusão de que 256 unidade por camada em 2 camadas escondidas e 100 dimensões oferecem resultados menos confusos e fora dos padrões planejados.

Encoding	One-hot	10 dim	100 dim	500 dim	1000 dim
512 units, 3 layers	2.86	2.30	2.58	1.71	1.77
512 units, 2 layers	2.75	2.05	2.59	1.93	1.74
256 units, 3 layers	3.12	2.07	1.70	1.70	1.71
256 units, 2 layers	3.10	2.59	1.61	1.63	1.72
192 units, 3 layers	3.42	2.10	1.83	1.73	1.74
192 units, 2 layers	4.20	1.86	1.80	1.78	1.74

Resultados

Quando usado para composição, as probabilidades dos acordes no dicionário serem próximos em uma sequência (Cp) podem ser armazenados através de algoritmos que baseando em apenas um acorde pode fazer uma sequência completa.

E assim que é feita a escolha o encoder passa aplicar esse acorde nas próximas batidas até uma troca. A ideia é imitar um músico que tem o controle de como os acordes se movem, e sabe como fazer as trocas e decidir como quer a música, mas também vai estar aberto a influência das decisões melódicas de outros músicos

Resultados

The image displays two musical staves, each containing six measures of music. The notes in the staves are represented by diagonal slashes, indicating that the focus is on the chord changes rather than the specific melodic lines. The first staff begins with a treble clef and a common time signature (C). Above the staff, the following chords are labeled: C⁷, Fmaj⁷, C⁷, C⁷, B^b7, E^b7, and A^b7. The second staff begins with a treble clef and a measure rest (7). Above the staff, the following chords are labeled: A^bmaj⁷, A^bmaj⁷, C⁷, B^b7, Dm⁷, D⁷, and D⁷.

Figura: Sequências de acordes geradas a partir de apenas um acorde (C⁷)

Integrando Melodia



Agentes em conjunto

Um agente de função de melodia foi introduzido para funcionar em cooperação com o agente de harmonia. Os dois agentes trabalham em um processo de co-agência e alternam entre os modos solo e de acompanhamento. O improvisador de melodia usa um banco de dados de frases baseadas em manipulações e permutações comuns de melodias humanas compostas e um modelo de harmonia baseado no conhecimento.

Os dois agentes usam um sistema de classificação para encontrar candidatos viáveis para a direção harmônica e melódica. Para desenvolver co-agência, as classificações são comparadas para selecionar qual agente assume um papel de liderança temporário.

Banco de Frases

Uma frase é uma sequência das 12 notas que formam uma melodia dentro de 1 a 4 barras. Cada melodia é revertida, invertida, aumentada e diminuída, seja separadamente e/ou em combinações.

Para cada frase(P), cada nota musical (P_i) tem um valor melódico(M_r) e classificação. Essas classificações se dividem em três categorias:

- Notas usáveis. (0.5)
- Notas do acorde. (1)
- Notas evitadas. (0)

Para mudanças de acorde é usada (0,25) para notas evitadas e usáveis. (b)

Chord type	Avoid notes	Chord tones
Major	m9, M4, M7	M3, M5
Dominant	m9, M4, M8	M3, M5, n7
Minor	m9	m3, M5

Estados de Liderança

O sistema possui dois estados, **liderança harmônica** e **liderança melódica**.

Na liderança harmônica, as progressões de acordes do estado inicial são geradas pelo agente de harmonia, selecionando o símbolo de acorde com o maior (Cp) para cada batida nas próximas duas barras. A média de (Cp) é registrada e usada como a classificação de acordes (Cr).

$$M_r = \frac{1}{L} \sum_{i=1}^L f(P_i) + b$$

O agente de melodia realiza então uma busca no banco de dados de melodia para encontrar uma linha melódica com Mr acima do valor desejado para tocar a progressão de acordes. Se nenhuma frase puder ser encontrada com uma classificação acima de um limiar desejado, o improvisador da melodia permanece até que uma frase seja aceita.

Estados de Liderança

No estado de liderança de melodia, o agente de melodia seleciona uma melodia para tocar e o agente de harmonia busca a saída da rede neural para cada batida seguinte do maior para o mais baixo C_p para encontrar um suporte harmônico para a melodia com M_r acima do valor desejado.

Quando na liderança de harmonia, uma mudança de estado é desencadeada quando uma linha melódica é encontrada com M_r maior que C_r para a próxima sequência ou uma linha melódica é encontrada com M_r acima do limiar estabelecido que ultrapassa dois compassos. No estado de chumbo da melodia, uma mudança de estado é desencadeada quando o improvisador harmônico propõe uma sequência de acordes com C_r maior que M_r para a linha melódica selecionada

Estados de Liderança

Ao juntar linhas melódicas ponta a ponta, temos uma fluidez na melodia como um todo, permitindo a criação de melodias maiores. Especificando a nota inicial para cada barra e adicionando outras regras tornam essa mudança mais natural.



Figura:

Mudanças de estado são desencadeadas por diferenças na classificação para a improvisação proposta de cada agente. Os estados de liderança da melodia também são desencadeados pelas melodias propostas que vão além das sequências de 2 barras propostas pelo agente de harmonia.

Resultados

The image displays a musical score for improvisation. It consists of two staves of music in 4/4 time, with a key signature of one flat (B-flat). The first staff features a melodic line with notes and rests, and a sequence of fingerings (1, 1, 1, 1, 0.5, 1, 0.5, 0.25, 1, 1, 0.5, 0.75, 0.5, 1, 1, 0.75, 1, 1, 1, 1, 1) written below it. Above the staff, four colored boxes indicate chord changes: a pink box for C7 (measures 1-4), a green box for Fmaj7 (measures 5-8), a red dashed box for C7 (measures 9-12), and a blue dashed box for C7 (measures 13-16). The second staff shows a bass line with notes and rests, and a sequence of fingerings (1, 0.5, 1, 1, 0.75, 1, 0.75, 1, 1, 0.75, 1, 1, 1, 0.75, 1) written below it. Above the staff, four blue dashed boxes indicate chord changes: C7 (measures 1-4), Bb7 (measures 5-8), Eb7 (measures 9-12), and Ab7 (measures 13-16).

Figura: Melodia improvisada em cima de uma progressão estática de acordes. Linhas sólidas representam a base para as variações da mesma cor.

Conclusão



Referências

HUTCHINGS, P. and McCORMACK, J. - “Using Autonomous Agents to Improvise Music Compositions in Real-Time”

[http://www-desir.lip6.fr/~briot/cours/unirio/Using Autonomous Agents to Improvise Mus.pdf](http://www-desir.lip6.fr/~briot/cours/unirio/Using_Autonomous_Agents_to_Improvise_Mus.pdf)

Quora - “What does thinking horizontally or vertically in music mean?”:

<https://www.quora.com/What-does-thinking-horizontally-or-vertically-in-music-mean>

MACHINE LEARNING DE SÉRIES TEMPORAIS – LSTM:

<https://www.monolitonimbus.com.br/machine-learning-de-series-temporais-lstm/>