

Comparação de Métodos de Cálculo de Similaridade de Características Multi-Valoradas: Uma Aplicação em Instrumentos MIDI

Marcel Lessa Oliveira¹, André Carvalho Silva¹, Aloísio de Menezes Vilas-Bôas¹, Hendrik Teixeira Macedo¹

Departamento de Computação - Universidade Federal de Sergipe
marcello@dcomp.ufs.br, andreacs@dcomp.ufs.br, aloisiomvb@dcomp.ufs.br,
hendrik@ufs.br

Abstract. This paper addresses the extraction of musical instruments in the MIDI format. Such a musical instrument is represented by a value which measures its relative participation within the music, concerning the total number of notes. In order to attest the representative power of a music through its musical instruments, an innovative method for similarity computation of multivalued features has been developed and applied to a corpus of musical archives. A second method proposed in the literature was equally applied and the results were analyzed. The results show that musical instruments can actually represent a music.

1 Introdução

A extração de características é um ramo de pesquisa bem desenvolvido. Particularmente, a extração musical já está bastante avançada, com diversas técnicas de extração sobre o sinal sonoro, como mostra [1]. Porém, essas características obtidas do sinal sonoro não são muito representativas. Um dos desafios é transformar um conjunto dessas características em outra característica mais representativa, como gênero.

Uma outra abordagem em relação a extração é utilizar outra representação da música, que não o seu sinal. Uma dessas representações é o formato de arquivo MIDI (*Musical Instrument Digital Interface*). O formato MIDI permite a representação de músicas tocadas por instrumentos equipados com a interface MIDI. Uma série de informações como frequência, velocidade e altura da nota são fornecidas pelo formato MIDI, bem como controles diversos independentes. Atualmente, também é possível utilizar softwares para criação de arquivos MIDI [2] [3].

A lista de aplicações nas quais é importante o uso de características extraídas é enorme. Um exemplo é a classificação automática de músicas por cantor [4] ou gênero [5]. Também é possível a classificação entre música ou fala com base no conteúdo [6].

Este trabalho apresenta um método para extração de instrumentos musicais de arquivos MIDI. Aliado a isso é proposto um método de cálculo de similaridade

de características multivaloradas, o qual é comparado a um método existente na literatura. Os resultados obtidos mostram que os dois métodos são equivalentes e que o método proposto é mais indicado para a similaridade entre instrumentos.

A seção 2 traz alguns trabalhos relacionados. A seção 3 apresenta o método desenvolvido para extração dos instrumentos de arquivos MIDI. O método proposto para cálculo de similaridade de características multivaloradas é apresentado na seção 4 enquanto que na seção 5 este método é comparado com um outro proposto na literatura. A seção 6 traz alguns testes realizados e os resultados obtidos. Finalmente, a conclusão do trabalho é apresentada na seção 7.

2 Trabalhos Relacionados

Uma abordagem interessante em extração musical é apresentada por [7]. Dixon extrai o tempo musical e as batidas, que configuram o ritmo de uma música. O processo de extração pode ser feito tanto no sinal sonoro, quanto em arquivos MIDI, sendo que cada caso apresenta algumas dificuldades técnicas. A abordagem é genérica de modo a contemplar qualquer tipo de música, e os resultados foram bastante positivos.

[8] utiliza arquivos MIDI de modo a criar um software didático musical. Os processos de criação das músicas eram feitos manualmente, tornando-os falhos e improdutivos. Com a automatização feita com o formato MIDI, o músico tem mais liberdade, de modo que sua execução musical é reconhecida pelo software. Um dos problemas com essa abordagem é a grande quantidade de ruído apresentada pelo violão MIDI.

A similaridade entre ritmos é abordada por [9], um tópico até então pouco explorado. Um dos problemas encontrados é que pequenas diferenças no tempo musical são comuns entre músicas. Os autores contornaram o problema utilizando um algoritmo de programação dinâmica chamado DTW (*Dynamic Time Warping*). A característica sonora que forneceu melhores resultados foi o *centróide espectral*. Deste modo eles conseguiram casar os padrões rítmicos de conjuntos de sons de percussão.

3 Extração de Instrumentos

Uma das contribuições deste trabalho é um método de extração de instrumentos musicais de arquivos MIDI. Geralmente a extração feita com o sinal sonoro trata de características de baixo nível desse sinal como taxa de passagem pelo zero, MFCC (*Mel-Frequency Cepstral Coefficients*), entre outros [10]. Entretanto, a extração do sinal sonoro por si só não é capaz de obter características consideradas de alto nível como ritmo, timbre ou até mesmo gênero. A classificação em gêneros é particularmente difícil, como cita [5], pois não existe uma fronteira exata entre os gêneros, tornando difícil até mesmo a classificação manual feita por especialistas.

O formato MIDI é composto de mensagens que são enviadas a um tocador MIDI. Essas mensagens são compostas de um tipo e o conteúdo da mensagem. A

maioria dos tipos de mensagens precisam ser direcionadas a um canal específico. O formato MIDI fornece 16 canais distintos para envio dessas mensagens. Uma das características do canal é que apenas um instrumento pode estar ativo em um certo momento. Exemplos de tipos de mensagens são: começar a tocar uma nota, terminar de tocar uma nota, alterar o volume do canal, alterar o instrumento do canal e alterar a velocidade de execução.

A figura 1 apresenta o modelo de uma mensagem MIDI padrão. O primeiro campo corresponde ao tempo decorrido desde a última mensagem. O segundo campo indica o tipo da mensagem e o canal ao qual ela é direcionada. Em algumas mensagens, esse campo representa apenas o tipo da mensagem. O terceiro campo contém o tamanho do conteúdo da mensagem e o quarto campo contém o conteúdo da mensagem em si.

1-4 bytes Δ -time	1 byte Tipo e canal	1-4 bytes Tamanho	<tamanho> bytes Conteúdo
-----------------------------	------------------------	----------------------	-----------------------------

Fig. 1. Formato de uma mensagem MIDI

A partir das informações presentes em um arquivo MIDI é possível obter características de alto nível. Uma dessas características é o conjunto de instrumentos que compõem uma música. Uma hipótese é que esta característica (instrumentos) seja representativa, uma vez que algumas pessoas são mais inclinadas a gostar de uma música se ela é tocada com instrumentos que elas apreciem. Para a característica consideramos a quantidade de notas por instrumento de uma música, de forma relativa.

A extração dos instrumentos de uma música é feita seguindo a ordem temporal da mesma, enquanto monitora-se os instrumentos dos canais. Cada nota enviada a um dos canais é contabilizada na contagem de notas do instrumento ativo daquele canal no momento do envio da nota. O resultado final é uma lista de instrumentos com uma quantidade de notas associada. Essa quantidade é apresentada de forma relativa, para ser usada posteriormente no cálculo de similaridade. O método de extração proposto é descrito resumidamente no algoritmo da figura 2.

O algoritmo percorre as mensagens da MIDI na ordem temporal. Se a mensagem for uma nota, a contagem de notas do instrumento ativo do canal ao qual foi enviada a mensagem é incrementada, assim como o total de notas lidas. Se a mensagem for uma mudança de instrumento, a referência ao instrumento ativo daquele canal é alterada. Por fim, todas as contagens de notas são divididas pelo total, de modo a obter o valor relativo de cada instrumento.

Algoritmo: Extração de instrumentos de um arquivo MIDI
Input: Um arquivo MIDI
Output: A lista de instrumentos com a participação relativa de cada um na música

```

(1)  while there is messages
(2)      read next message in time
(3)      case message type
(4)          note →
(5)              notes[instruments[channelInstruments[channel]]]++
(6)              totalNotes++
(7)          instrument →
(8)              channelInstruments[channel] = instrument
(9)  for i = 1 to instruments.size
(10)     instruments[i] = instruments[i] / totalNotes
(11)  return instruments

```

Fig. 2. Pseudo-código do algoritmo proposto de extração

4 Método de cálculo de similaridade de características multivaloradas

A comparação entre características de objetos é um processo natural, que serve para identificá-los. Quanto mais características similares dois objetos possuíam, mais eles se enquadram num mesmo tipo de objeto. Logo, após a extração de características é necessário comparar as características de objetos diferentes. Porém, nem todas as características possuem métodos diretos para serem comparadas, como é o caso dos instrumentos.

Essa comparação, feita através do cálculo de similaridade entre os instrumentos de duas músicas, ajudará de forma indireta na verificação da hipótese levantada na seção anterior. A idéia é que se uma pessoa gostar de músicas similares em relação a instrumentos, então os instrumentos são uma característica representativa nas músicas.

Segundo Byron [11], características ou variáveis multivaloradas “são conjuntos de categorias ou de números, intervalos ou distribuições de probabilidade associadas a regras e taxonomias”. Os instrumentos extraídos da MIDI se enquadram nesta definição, pois são um conjunto de números (quantidade de notas) associados a uma taxonomia, que é a separação das notas nos diversos instrumentos. O método de cálculo de similaridade foi primeiramente desenvolvido tendo em foco a similaridade entre instrumentos, mas pode facilmente ser utilizado para outras características que se comportem de modo similar.

O método proposto tem algumas características que devem ser notadas:

1. O método é uma operação comutativa, tal que $S(A, B) = S(B, A)$
2. O método foi planejado considerando que as características são apresentadas com valores relativos
3. O método também leva em consideração possíveis agrupamentos de instrumentos, por exemplo, instrumentos de sopro

O método consiste na paridade dos instrumentos de duas músicas distintas, com a posterior seleção desses pares. Os pares, chamados de relações, são o conjunto criado a partir do produto cartesiano entre os instrumentos das músicas. No caso, o conjunto R de relações do tipo $r(i_A, i_B)$ é o conjunto:

$$R = I_A \times I_B \quad (1)$$

onde I_X representa o conjunto dos instrumentos de uma determinada música X e i_X é um elemento deste conjunto.

Cada relação tem um valor $r.v$ que é dado por:

$$\forall r(i_A, i_B) \in R, r.v = \frac{\min(i_A, i_B)}{\max(i_A, i_B)} * f \quad (2)$$

Na equação 2, f corresponde ao fator de grupo, que tem um valor variante de 0 a 1. Com instrumentos diferentes de um mesmo grupo, essa variável pode receber um valor intermediário. Para instrumentos iguais f tem valor 1, e para instrumentos diferentes que não do mesmo grupo, o valor é 0.

A similaridade é calculada incrementalmente seguindo as fórmulas 3 e 4:

$$S(A, B) = \frac{S_1 + S_2 + \dots + S_n}{n} \quad (3)$$

$$S_i = \max(R_i) \quad (4)$$

onde $\max(R_i)$ é a relação cujo $r.v$ é máximo no conjunto R_i . A cada iteração i a relação de maior valor é selecionada. Então o novo conjunto R_{i+1} é criado de acordo com as equações 5 e 6.

$$R_{i+1} = R_i - E_i \quad (5)$$

$$E_i = \{r(x, y) \in R_i | \exists z, w, \max(R_i) = r(z, w), x = z \vee y = w\} \quad (6)$$

O conjunto E_i representa todas as relações do conjunto R_i que compartilhem pelo menos 1 instrumento com $\max(R_i)$. Esse conjunto será subtraído do conjunto R_i , formando o conjunto R_{i+1} , a partir do qual o processo se repete iterativamente. Ao final de todas as iterações é calculada a média dos valores das relações máximas selecionadas ao longo do processo. O cálculo inicia-se com $R_1 = R$ e termina quando o conjunto R_n é tal que $R_{n+1} = \phi$. O processo pode ser visto visualmente na figura 3

5 Comparação entre métodos de cálculo de similaridade

Um dos métodos mais relevantes na literatura para cálculo de similaridade de características multivaloradas é utilizado por Byron [11] e foi desenvolvido por Ichino e Yaguchi [12]. O método em questão, entretanto, não atende plenamente

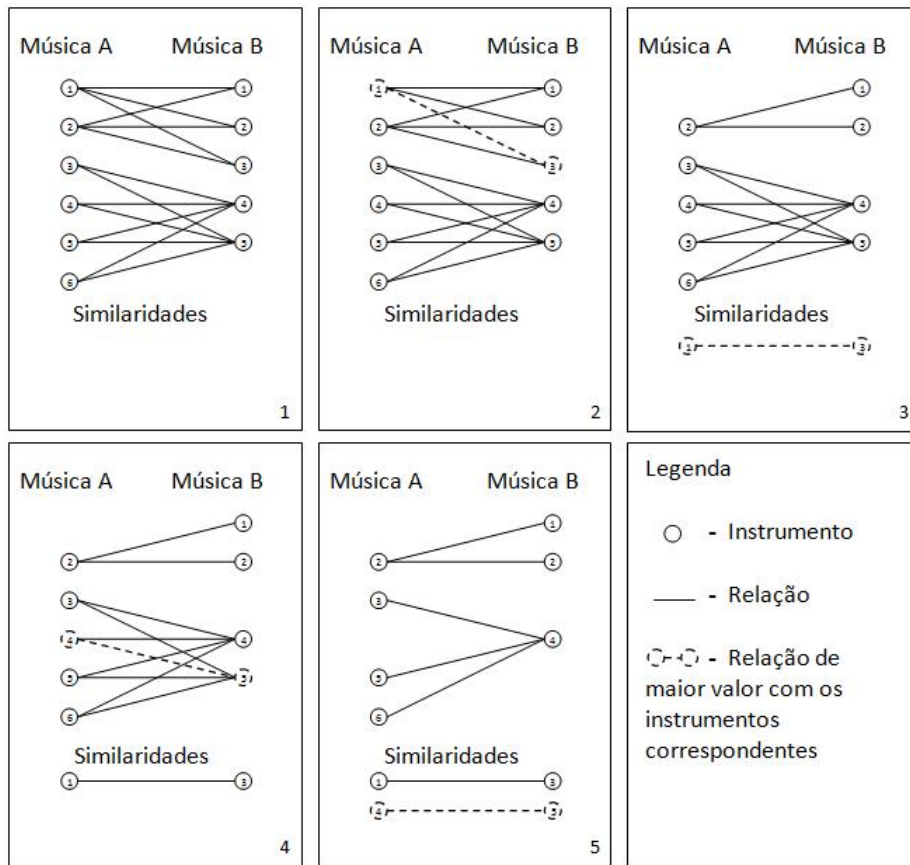


Fig. 3. Método de cálculo de similaridade proposto; 1: início, 2: seleção da relação de maior valor, 3: remoção das relações com mesmo instrumento, 4 e 5: repetição dos passos 2 e 3

às necessidades específicas da característica explorada. Por esta razão, este trabalho propõe um novo método para o cálculo da similaridade, e que é descrito logo a seguir.

O método proposto calcula a dissimilaridade entre objetos e é composto por duas componentes: uma chamada *livre de contexto*, usada para características monovaloradas, e outra *dependente de contexto*, para características multivaloradas. O método é descrito pelas equações 7, 8 e 9, mas vamos explorar apenas a 9, que corresponde à componente dependente de contexto.

$$\phi(x, u) = \sum_{i=1}^p \phi_{cf}(x_i, u_i) + \phi_{cd}(x_i, u_i) \quad (7)$$

$$\phi_{cf}(x_i, u_i) = \frac{|\overline{X}_i \cap \overline{U}_i \cap (X_i \oplus U_i)|}{|X_i \oplus U_i|} \quad (8)$$

$$\phi_{cd}(x_i, u_i) = \frac{1}{2} \left[\sum_{k/x_k \in X_i \cap \overline{U}_i} w_k + \sum_{m/u_m \in \overline{X}_i \cap U_i} W_m \right] \quad (9)$$

Essa componente dependente de contexto soma os valores dos instrumentos que só aparecem de um dos lados, dividindo tudo por 2. Como este método calcula a dissimilaridade, calculamos a similaridade utilizando os instrumentos que se repetem em ambos os lados da comparação. Isso equivale a somar as médias dos instrumentos comuns às duas músicas.

Ambos os métodos são utilizados para calcular (dis)similaridade entre características, porém o método proposto neste artigo e o proposto por Ichino [12] apresentam algumas diferenças. Primeiramente, o método de Ichino não dá suporte a grupos de instrumentos, apenas instrumentos iguais contam para a similaridade. Em um caso extremo, uma música que contenha somente sons de guitarra não seria similar a outra que tenha apenas violão, apesar de ambos os instrumentos serem do mesmo grupo.

Outra diferença entre os métodos é o cálculo de similaridade entre os instrumentos. O método proposto utiliza uma razão, enquanto o método de Ichino usa uma média. A tabela 1 mostra como essa diferença influencia na similaridade. Em termos práticos, duas músicas onde o mesmo instrumento seja tocado poucas vezes teriam uma similaridade baixa pelo método de Ichino, dado a baixa participação do instrumento em questão. Já com o método proposto, a similaridade deste instrumento seria alta, pois a participação em ambas as músicas seriam próximas.

Table 1. Similaridades utilizando média e razão

Porcentagem		Similaridade	
Música A	Música B	Média	Razão
Alta	Alta	Alta	Alta
Alta	Baixa	Média	Baixa
Baixa	Baixa	Baixa	Alta

6 Testes e resultados

Foram realizados alguns testes para aferir a eficácia de ambos os métodos apresentados. A hipótese é que o método proposto apresentará melhores resultados, pois foi planejado com foco nos instrumentos. No primeiro teste foi utilizado o algoritmo de agrupamento *k-means* onde a distância entre as músicas era a similaridade entre os instrumentos. Foi utilizado nesse teste um *corpus* composto

de 274 músicas de diversos gêneros [13]. Os grupos foram associados aos gêneros dominantes da população. Os dois métodos comportaram-se de maneiras diferentes: o método de Ichino agrupou a maior parte das músicas em um único grupo e o método proposto distribuiu mais as músicas. O erro médio e o desvio-padrão médio (em unidades) para o método proposto foram de 74% e 45,6 e o método de Ichino obteve 67% e 48,6. O erro do método proposto foi maior, mas em compensação seu desvio-padrão foi menor.

Em outro teste, foi utilizado um sistema de recomendação baseado em conteúdo, desenvolvido paralelamente. Um sistema de recomendação baseado em conteúdo analisa as músicas avaliadas por um usuário e, de acordo com a similaridade entre as músicas, atribui notas a músicas ainda não avaliadas pelo usuário. O sistema de recomendação utilizava três características ao recomendar uma música: o gênero, o artista e os instrumentos. Assim, para um mesmo grupo de usuários, foi calculado os erros e desvios-padrão médios de classificação em cinco configurações de características: duas com gênero, artista e instrumentos, variando o método de comparação de instrumentos entre o proposto e o de Ichino; duas apenas com instrumentos, novamente variando o método de comparação; e uma com gênero e artista. Este teste também verifica a hipótese apresentada no capítulo 2. O resultados encontram-se dispostos na tabela 2.

Table 2. Resultados dos testes

Teste	Erro médio	Desvio-padrão
Agrupamento (proposto)	74%	45,6
Agrupamento (Ichino)	67%	48,6
Recomendação (completo (proposto))	2,09%	1,30%
Recomendação (completo (Ichino))	2,16%	1,32%
Recomendação (instrumentos (proposto))	2,25%	1,34%
Recomendação (instrumentos (Ichino))	2,31%	1,36%
Recomendação (gênero, artista)	4,48%	2,26%

A tabela mostra que o método proposto apresenta resultados ligeiramente melhores em ambos os parâmetros. Vemos também que a inclusão de instrumentos no rol de características do sistema de recomendação melhora consideravelmente os resultados.

7 Conclusão

Existem muitos tipos de aplicações onde é possível usar extração de características. Um dos problemas na extração de características musicais é a dificuldade em transformar as características de baixo nível obtidas a partir do sinal sonoro em características mais representativas, como por exemplo gênero. Uma outra abordagem é obter as características de uma representação em alto nível, como o formato MIDI. Porém, as características possíveis de serem extraídas são limitadas.

Uma característica que foi possível extrair do formato MIDI foram os instrumentos, tratados neste trabalho. Junto com sua extração foi desenvolvido um método de comparação entre características, dado que os métodos existentes não atendiam às necessidades do trabalho. O método proposto foi comparado com um método encontrado na literatura. O novo método mostrou-se equivalente ao método clássico em um teste controlado, mas num teste de aplicação real, o método proposto foi ligeiramente superior ao da literatura, atendendo às expectativas.

Os resultados dos testes mostraram também que não é possível classificar músicas em gêneros, tendo como base apenas instrumentos, dado o alto erro de classificação. No entanto, mais testes precisam ser feitos para verificar a influência dos instrumentos nos gêneros. Mas a influência dos instrumentos na recomendação é um forte indício da importância desta característica no gosto musical das pessoas, confirmando a hipótese de que os instrumentos são uma característica representativa nas músicas.

References

1. Pickens, J.: A survey of feature selection techniques for music information retrieval. Technical report, CIIR - Center for Intelligent Information Retrieval (2001)
2. Image-Line: FL studio. Disponível em: <http://flstudio.image-line.com/index.html> (2008)
3. Cakewalk: SONAR 8. Disponível em: <http://www.cakewalk.com/Products/SONAR/English/default.asp> (2009)
4. Liu, C.C., Huang, C.S.: A singer identification technique for content-based classification of mp3 music objects. In: In Proceedings of International Conference on Information and Knowledge Management. (2002) 438–445
5. Li, T., Ogihara, M., Li, Q.: A comparative study on content-based music genre classification. Special Interest Group on Information Retrieval Conference (2003)
6. Khan, M.K.S., Al-Khatib, W.G., Moinuddin, M.: Automatic classification of speech and music using neural networks. International Workshop on Multimedia Databases (2004)
7. Dixon, S.: Automatic extraction of tempo and beat from expressive performances. Journal of New Music Research (2001)
8. de Sá, A.S.A.: Extração de conhecimento musical a partir de execuções violinísticas. Universidade Federal de Pernambuco (2006)
9. Paulus, J., Klapuri, A.: Measuring the similarity of rhythmic patterns. In: Proceedings of the International Symposium on Music Information Retrieval. (2002) 150–156
10. McKinney, M.F., Breebaart, J.: Features for audio and music classification. In: Proceedings of the International Symposium on Music Information Retrieval. (2003) 151–158
11. Bezerra, B.L.D.: Estudo de algoritmos de filtragem de informação baseados em conteúdo. Universidade Federal de Pernambuco (2002)
12. Ichino, M., Yaguchi, H.: Generalized minkowsky metrics for mixed feature type data analysis. In: IEEE Transactions system, Man and Cybernetics. Volume 24. (1994) 698–708

13. Basili, R., Serafini, A., Stellato, A.: Classification of musical genre: a machine learning approach. In: Proceedings of the International Symposium on Music Information Retrieval. (2004)