



ESTUDO DO IMPACTO DA SEGMENTAÇÃO SEMÂNTICA NA CLASSIFICAÇÃO DE MÚSICAS NO CENÁRIO MULTIRRÓTULO

Vinícius Menossi (PIBIC/CNPq/FA/UEM), Yandre Maldonado e Gomes da Costa (Orientador), Juliano Henrique Foleiss(Coorientador) e-mails: ra108840@uem.br, yandre@din.uem.br, julianofoleiss@gmail.com.

Universidade Estadual de Maringá / Centro de Tecnologia / Maringá, PR.

10303006 METODOLOGIA E TÉCNICAS DA COMPUTAÇÃO

Palavras-chave: classificação de gêneros de musicais, classificação multirrótulo, segmentação semântica de áudio.

Resumo:

Considerando as crescentes pesquisas na classificação de músicas de forma automatizada, esse trabalho busca investigar os impactos da segmentação dinâmica na extração de características no cenário multirrótulo através do algoritmo KMEANSC.

Introdução

A classificação automática de músicas é uma tarefa que vem há tempos sendo investigada pela comunidade científica, esse interesse se deve principalmente ao grande aumento no consumo dos formatos digitais destas mídias. Inicialmente, muitos trabalhos na literatura científica se voltaram a investigar a classificação de gêneros musicais.

Os gêneros musicais são rótulos criados por humanos - especialistas e amadores - para caracterizar estilos de música. Esses rótulos podem derivar de diversas características presentes na música como a instrumentalização utilizada, estrutura rítmica e conteúdo harmônico. O cenário onde cada música é associada a um único gênero musical dentre vários possíveis, é descrito como um problema de classificação multiclasse.

Em estudos posteriores se investigou a hipótese de cada título musical estar associado a mais de um único gênero musical. Esse cenário onde vários rótulos podem ser associados a um único título é caracterizado como um problema de classificação multirrótulo.

Recentemente houve um aumento no número de pesquisas realizadas utilizando outros rótulos associados a títulos musicais, que não necessariamente correspondem a gêneros musicais, como por exemplo a classificação de sentimentos. Esses rótulos, muitas vezes são obtidos coletando-se termos frequentemente utilizados por usuários de rádios *online* ou serviços de *streaming* e associados aos títulos musicais.

Em todos os cenários apresentados de classificação de músicas aqui descritos, um problema vigente é como realizar a amostragem do sinal da música adequadamente, para realizar a classificação, de forma que sejam utilizadas partes efetivamente relevantes da música para o problema de classificação. A forma com que o sinal de áudio é amostrado impacta diretamente a classificação, podendo ser uma tarefa decisiva nos resultados. O uso de pequenos trechos e recortes do sinal tornou-se uma abordagem tradicional visando a diminuição do tamanho do processamento a ser realizado nas etapas de extração de características, treinamento e teste. Em geral essa amostragem era realizada escolhendo um ou mais trechos da música, com seus tamanhos sendo pré-definidos de maneira fixa para toda e qualquer música da base a ser analisada.

Foleis e Tavares (2020) apresentaram o KMEANSC, um algoritmo que pode selecionar automaticamente os trechos mais promissores de um sinal de áudio na busca da melhoria na efetividade do desenvolvimento de classificadores automáticos. O KMEANSC pode selecionar quantidades de segmentos distintos para músicas diferentes, e o tamanhos desses trechos selecionados também pode variar, o algoritmo trata cada música como um caso particular, e as propriedades específicas na definição dos um ou mais trechos a serem selecionados. O trabalho dos autores mostra que esse algoritmo é capaz de capturar propriedades diferentes para trechos distintos da música, que em hipótese poderiam ser associadas a rótulos variados previstos em bases de música criadas na perspectiva multirrótulo.

Neste trabalho é feita investigação, pelo uso do KMEANSC, da hipótese de que partes diferentes da música, com propriedades de ritmo, timbre e harmonia distintas, podem capturar de forma mais abrangente diferentes rótulos eventualmente associados a um título musical, o que poderia melhorar os resultados da classificação de músicas no cenário multirrótulo.

Materiais e Métodos

Bases escolhidas

A princípio foi realizado um teste de comparação para uma classificação de multiclasse no cenário monorrótulo. A base de dados escolhida foi a GTZAN, contendo 1000 trechos de música com a duração de 30s, contendo 10 gêneros e 100 músicas para cada gênero. Para o teste de classificação no cenário multirrótulo foi escolhida a base CAL500, constituída de 502 músicas distribuídas em 36 gêneros, cada música presente na base pode assumir 1 a 36 desses gêneros. Na proposta inicial do projeto havia sido selecionada a base Music4all, porém não existe uma versão da base disponível que ofereça as músicas com o seu conteúdo integral, o que é indispensável para a avaliação da estratégia de segmentação aqui investigada (KMEANSC).

Segmentação do sinal

Para base GTZAN não houve a segmentação estática, pois os trechos disponibilizados eram de 30 segundos de duração, na base CAL500 foram retirados 3 trechos de cada música, o primeiro do segundo 16 ao segundo 25, o segundo, sendo a duração da música constituída de n segundos, e o meio dessa $m = n/2$, o

segundo trecho se tratou dos segundo $m-5$ ao segundo $m+5$, já o terceiro trecho do intervalo do segundo $n-25$ ao segundo $n-15$. Ilustrado na figura 1.



figura 1: segmentação do sinal

Já o algoritmo KMEANSC segmenta de forma dinâmica o sinal do áudio produzindo tamanhos variados de trechos do áudio.

Extração das Características

A classificação de sinais de áudio estão relacionadas a diversas características musicais como: melodia, harmonia, ritmo, timbre e localização espacial. As características consideradas de baixo nível na representação musical podem ser definidas como características relacionadas ao timbre, variação de frequência, e rítmicas, Tzanetakis e Cook(2002). O timbre corresponde às distintas características entre dois sons com a mesma frequência sejam percebidos de forma diferente, o formato da onda. A variação da frequência está relacionada a harmonia e melodia do som. O ritmo está relacionado aos aspectos temporais da música.

Utilizando a biblioteca *librosa* e o framework *marsyas*, foram extraídas as seguintes características, tempo médio estimado da música em BPM's, zero crossing, 20 coeficientes mel-cepstrais mffcs's, e 12 chroma's com a transformada rápida de Fourier, após a extração foram retirados a média e o desvio padrão de cada uma dessas medidas totalizando um vetor de 68 posições. Essas características foram retiradas dos sinais segmentados e também das músicas inteiras da base CAL500.

Classificadores

Os classificadores utilizados nesse trabalho através da biblioteca *sklearn*, foram o Support Vector Machines(SVM) no teste com classificação multiclasse e monorrótulo, já o classificador Random Forest foi utilizado na classificação multirrótulo, com uma cadeia de classificação, útil no cenário em que existe correlação entre os rótulos, como por exemplo raramente ocorre a presença dos rótulos do gênero rap e jazz simultaneamente.

A validação do treinamento foi realizada utilizando a função de validação cruzada do *sklearn*, utilizando 3 subconjuntos dos dados.



Resultados e Discussão

No primeiro teste realizado no cenário multiclasse, com o classificador SVM, na base GTZAN com 1000 trechos de músicas e 10 gêneros, a abordagem com segmentação estática obteve um score f1 médio de 73%, já a segmentação dinâmica através do KMEANSC obteve um f1 score médio de 79%.

Já no teste realizado no cenário de classificação multirrótulo, com o classificador *Random Forest* na base CAL500 com 500 músicas e 36 gêneros possíveis, a abordagem sem segmentação e apenas com o sinal inteiro obteve um score f1 de 9%, e uma precisão de 27%, a segmentação estática obteve o score f1 de 11%, e uma precisão de 31% enquanto através da segmentação dinâmica pelo algoritmo KMEANSC foi possível obter um score f1 de 19% e uma precisão de 41%.

Conclusões

Com a análise dos modelos treinados é possível perceber uma melhora na performance da classificação quando utilizada a segmentação do semântica, tanto no cenário de classificação multiclasse e multirrótulo. Apesar disso, no cenário multirrótulo obteve resultados baixos mesmo na segmentação dinâmica do sinal na base CAL500, uma possível explicação seria o tamanho pequeno da base .

Agradecimentos

À Fundação Araucária e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, por me possibilitarem a bolsa para que esse projeto fosse realizado. À Universidade Estadual de Maringá que possibilitou a execução deste trabalho através do Programa Institucional de Bolsas de Iniciação Científica - PIBIC. Ao prof. Yandre Maldonado Gomes da Costa, orientador deste trabalho, ao prof. Juliano Henrique Foreis, coorientador deste trabalho, que também disponibilizou o código fonte do algoritmo KMEANSC que foi crucial para o desenvolvimento desse tema.

Referências

- Y. M. G. Costa, L. E. S. Oliveira, A. L. Koerich, F. Gouyon, J. G. Martins. Music genre classification using LBP textural features. *Signal Processing*, 92(11):2723-2737, 2012.
- J. H. Foleis, T. F. Tavares. Texture selection for automatic music genre classification. *Applied Soft Computing*.
- G. Tzanetakis e P. Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293-302, 2002.