

Kent Academic Repository

Full text document (pdf)

Citation for published version

Silla Jr, Carlos N. and Kaestner, Celso A.A. and Koerich, Alessandro L. (2007) The Latin Music Database: Uma Base de Dados Para a Classificacao Automatica de Generos Musicais. In: 11th Brazilian Symposium on Computer Music. pp. 167-174.

DOI

Link to record in KAR

<https://kar.kent.ac.uk/24092/>

Document Version

UNSPECIFIED

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

The Latin Music Database: Uma Base de Dados Para a Classificação Automática de Gêneros Musicais

Carlos N. Silla Jr.¹, Celso A. A. Kaestner², Alessandro L. Koerich¹

¹Pontifícia Universidade Católica do Paraná (PUC-PR)
Programa de Pós-Graduação em Informática Aplicada (PPGIA)

²Universidade Tecnológica Federal do Paraná (UTFPR)
Curitiba - PR - Brasil

silla@ppgia.pucpr.br, kaestner@utfpr.edu.br, alekoe@ppgia.pucpr.br

Abstract. *In this paper we present the Latin Music Database, a novel database of Latin musical recordings which was developed for automatic music genre classification but can also be used to other tasks related to music information retrieval (MIR) research. The method for assigning genres to the musical recordings is a novel one and it is based on human perception. Furthermore, the underlying framework allows that the database can be easily expanded and have all the features desired by the research community of audio information retrieval.*

Resumo. *Este artigo descreve a Latin Music Database, uma base de dados nova e inédita, que foi originalmente criada para experimentos relacionados a classificação automática de gêneros musicais, mas que pode ser utilizada para diversas tarefas relacionadas a recuperação automática de informações musicais (MIR). O método para atribuição de gêneros musicais também é inédito e é baseado na percepção humana. Além disso, é apresentado o framework desenvolvido que garante que a base de dados seja facilmente expansível e tenha todas as características desejadas atualmente pela comunidade científica da área de recuperação automática de informações musicais.*

1. Introdução

A área de recuperação automática de informações musicais tem recebido cada vez mais atenção nos últimos anos. Contudo apesar da grande quantidade de métodos e aplicações que surgem, o número de bases de dados públicas e confiáveis para realizar experimentos ainda é pequeno. Considerando os trabalhos da área de classificação automática de gêneros musicais, que foram impulsionados pelo trabalho de Tzanetakis & Cook [9] que apresentaram a tarefa de classificação automática de gêneros musicais como sendo um problema de reconhecimento de padrões, existe um déficit muito grande em bases de dados que contenham um número suficiente de exemplos e que tenham sido rotuladas de forma a permitir que a base seja utilizada em outros experimentos.

Essa questão só foi recentemente abordada no trabalho de McKay et al. [7] onde são descritas as características desejadas na construção de novas bases de dados musicais, de forma que elas sejam realmente aproveitadas pela comunidade científica ao invés de terem sem uso limitado a dois ou três experimentos, como normalmente acontece. Neste mesmo trabalho de McKay et al. [7] é apresentada a base de dados CODAICH, que possui 20.894 músicas no formato MP3 de 1.941 artistas. De acordo com este trabalho, as características desejadas em novas bases de dados são:

1. Os dados devem ser distribuídos para os pesquisadores de forma livre e legal (em função das questões de direitos autorais).
2. A base de dados deve conter diferentes tipos de música.
3. A base de dados deve incluir um grande número de músicas.
4. A base de dados deve incluir uma quantidade significativa de músicas comerciais, pois os usuários finais estão principalmente interessados em músicas produzidas profissionalmente.
5. Cada música deve ser rotulada com a maior diversidade de campos possíveis, de forma a permitir que a base de dados seja utilizada em uma ampla gama de experimentos.
6. Idealmente, seria possível atribuir rótulos não apenas para a música como um todo, mas também para os seus segmentos.
7. Os rótulos de campos subjetivos como gênero ou humor (*mood*) devem permitir uma ampla variedade de rótulos possíveis.
8. A atribuição dos rótulos deve ser correta, completa e consistente.
9. Deve ser possível atribuir múltiplos rótulos independentes para um único campo, para que seja possível, por exemplo, classificar uma música como swing e blues.
10. Meta informações devem ser fornecidas aos usuários em formatos que sejam fáceis tanto de lidar manualmente quanto usando ferramentas automáticas.
11. Músicas inteiras devem estar acessíveis, mesmo que de forma indireta, e não apenas trechos das músicas.
12. Considerando que diferentes métodos de compressão podem influenciar os valores das características extraídas, o(s) formato(s) de áudio mais comuns ao público devem ser utilizados de forma que os experimentos realizados possam refletir condições realistas.

Além da CODAICH database, as únicas bases de dados atualmente disponíveis para a tarefa de classificação automática de gêneros musicais são: a base de dados GTZAN, desenvolvida no trabalho de Tzanetakis & Cook [9], que possui 1.000 músicas de 10 gêneros distintos (Blues, Clássica, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae e Rock), sendo 100 músicas de cada gênero; e a base de dados desenvolvida no trabalho de Homburg et al. [3], que possui 1.886 músicas obtidas a partir do site Garageband. A base está dividida em 9 gêneros sendo: Blues (120); Electronic (113); Jazz (319); Pop (116); Rap/HipHop (300); Rock (504); Folk/Country (222); Alternative (145); Funk/Soul (47).

Porém essas duas bases possuem sérias limitações: na primeira estão disponíveis apenas os primeiros 30 segundos de cada música no formato de áudio PCM. Na segunda estão disponíveis apenas 10 segundos extraídos de segmentos aleatórios de cada música. Com exceção dessas duas bases, as demais bases utilizadas na literatura normalmente possuem poucas músicas, e os gêneros utilizados são normalmente os mesmos (Rock, Clássica) e normalmente os gêneros são disjuntos, ou seja, não existem trabalhos com subgêneros realmente próximos como House e Trance.

Neste trabalho é apresentada a Latin Music Database e como todos os requisitos desejáveis para a construção de uma nova base de dados foram abordados no seu desenvolvimento. O restante deste trabalho está organizado da seguinte maneira: na seção 2 são apresentados os detalhes referentes a Latin Music Database e o framework desenvolvido para trabalhar com a mesma; e na seção 3 são apresentadas as conclusões deste trabalho.

2. Aspectos Referentes a Latin Music Database e seu Framework

Como visto na seção anterior as bases de dados existentes para realizar a tarefa de classificação automática de gêneros musicais possuem uma série de problemas e/ou

limitações. Dessa forma existe a necessidade do desenvolvimento de uma nova base de dados para a tarefa, porém no intuito de desenvolver uma base de dados confiável e que possa ser utilizada como *ground-truth* em outras pesquisas da área (não limitadas apenas a classificação automática de gêneros musicais) uma série de cuidados foram tomados. Essas precauções são apresentadas a seguir.

2.1. Criação e Manutenção da Base de Dados

Tendo em vista as limitações das bases desenvolvidas nos trabalhos anteriores na área de classificação automática de gêneros musicais [2] [4], surgiu a necessidade do desenvolvimento de uma nova base de dados para a tarefa. Porém considerando o esforço humano necessário para fazer a atribuição manual de gêneros às músicas, e também que uma base desenvolvida com cuidado poderia ser utilizada em outras tarefas além da classificação automática de gêneros musicais, foi necessário planejar como seria realizada a atribuição dos gêneros e o armazenamento, acesso e recuperação dessas informações.

Antes de iniciar o processo de aquisição, classificação e armazenamento das músicas foi definido que seriam adquiridas pelo menos 3.000 músicas de 10 gêneros distintos de forma a poder fazer uma contribuição real para a área, visto que até então a base de dados mais abrangente (GTZAN) era composta por 1.000 músicas (limitadas a apenas os primeiros trinta segundos) de 10 gêneros.

2.2. O Processo de Atribuição de Gêneros Musicais

Neste trabalho o processo utilizado para atribuir um gênero a cada música, é baseado na percepção humana de como cada música é dançada. Para realizar esse processo foram consultados dois profissionais com mais de dez anos de experiência no ensino de dança e danças de salão. Esses profissionais fizeram uma primeira seleção das músicas que eles julgavam pertinentes a um determinado gênero de acordo com a forma que este era dançado e o primeiro autor deste trabalho verificou cada uma das músicas inicialmente selecionadas para evitar que equívocos fossem cometidos devido ao desgaste produzido pelo esforço humano necessário para realizar a tarefa. Em média foram classificadas 300 músicas por mês, sendo que o processo total para a criação da base de dados demorou um ano.

Como resultado desse esforço, foi desenvolvida a *Latin Music Database* que conta com 3.160 músicas de 10 gêneros musicais no formato MP3. Os gêneros musicais disponíveis na base e seus respectivos números de músicas são: Tango (404); Salsa (303); Forró (315); Axé (304); Bachata (308); Bolero (302); Merengue (307); Gaúcha (306); Sertaneja (310); Pagode (301). No total a base possui 543 artistas diferentes.

É importante ressaltar que na base desenvolvida foi utilizado esse protocolo de inspeção humana utilizando uma perspectiva da percepção humana de como as músicas são dançadas, pois ao contrário do que foi sugerido no trabalho de Aucouturier & Pachet [1] de utilizar CDs de coleções completas, no caso dos ritmos latinos essa abordagem se mostrou ineficiente. Por exemplo, no caso da coletânea de quatro CDs (*Los 100 Mayores Exitos De La Musica Salsa*) apenas metade (50 das 100) das músicas podem ser classificadas como Salsa, as demais músicas dessa coletânea são de outros gêneros musicais como Merengue, Lambada, Zouk e até mesmo Samba. Outra opção teria sido basear a classificação de todas as trilhas de um determinado álbum de acordo com o perfil do artista, dessa forma todas as músicas de Carlos Gardel seriam classificadas como Tango, porém é importante ressaltar, que de todas as suas mais de 500 composições apenas cerca de 400 são Tangos, dessa forma introduzindo ruído desnecessário na base. Por esse motivo todas as músicas utilizadas nesta base foram avaliadas manualmente e somente aquelas que realmente pertencem aos gêneros em questão foram rotuladas como sendo desses

gêneros. E mesmo no caso de outros artistas de um determinado gênero, como Salsa, muito dificilmente todas as trilhas de seus álbuns são apenas Salsas.

Ao longo do processo de criação da base foi observado que normalmente cerca de uma a três músicas não são do gênero principal do perfil do artista.

2.3. Armazenamento, Acesso e Recuperação das Músicas

Além da aquisição das músicas e suas respectivas atribuições de gênero, para o desenvolvimento da base e sua ampla utilização em outras tarefas, várias reflexões foram realizadas no sentido de: criar uma base que possa ser facilmente utilizada para outras tarefas; permitir total reprodutibilidade dos experimentos realizados; evitar duplicidade nas músicas cadastradas; facilitar o registro de novas músicas e/ou novos gêneros. As reflexões realizadas no desenvolvimento do framework refletem os requisitos desejados apresentados no trabalho de McKay et al. [7] e listados anteriormente. Dessa forma, tendo em mente esses vários requisitos, nesta seção são apresentadas às soluções adotadas para atingir esses objetivos. O processo de armazenamento de uma nova música na base ocorre da seguinte forma:

1. Atribuição de um gênero a música em questão seguindo o procedimento descrito na Subseção 2.2;
2. Inspeção manual do rótulo ID3 da música para verificar se os campos estão preenchidos corretamente e também de corrigi-los/adaptá-los a um padrão simples que consiste na padronização dos nomes e no uso do caractere especial & para indicar o nome de mais de um artista na mesma música. Os campos obrigatórios para cadastrar uma nova música são o *Artista* e o *Título* da música. A razão para essa abordagem é simples, mesmo que apenas uma pessoa esteja trabalhando no cadastro de músicas na base de dados, eventualmente álbuns do mesmo artista vão conter trilhas com músicas presentes em outros álbuns, como por exemplo, no caso de um álbum com os maiores sucessos de um artista. Dessa forma, esse procedimento permite evitar duplicidade de músicas interpretadas pelo mesmo artista na base. Esse controle de duplicidade é realizado no sistema quando uma nova música vai ser cadastrada.
3. Cadastramento da música no sistema. Nessa etapa o sistema obtém os dados da música, verifica se não há duplicidade, atribui um código identificador para a música, associa essa música ao gênero pré-determinado e cria uma cópia da música. A informação de qual o gênero da música é armazenada no banco de dados pois como visto anteriormente o campo *Genre* dos rótulos ID3 não é confiável. Além disso, no caso de trabalhos futuros onde seja necessário o uso de alguma hierarquia, essa modificação pode ser incorporada facilmente ao sistema. No momento do cadastramento o sistema gera uma cópia da música cadastrada em um diretório pré-determinado seguindo a seguinte convenção:

DIRETORIO_GENERO\ARTISTA - TITULO - ALBUM - TRACK.MP3

onde DIRETORIO_GENERO é um diretório com o nome do gênero associado a música, e ARTISTA, TITULO, ALBUM e TRACK são informações obtidas do rótulo ID3 da música no momento em que ela é cadastrada.

O acesso a base de dados pode ser feito de forma convencional através do sistema de arquivos do sistema operacional, pois como visto, o sistema utiliza uma estrutura de arquivos e algumas regras de convenção simples para cadastrar as músicas.

As principais tabelas no sistema são: MUSIC, MUSIC_INFO e GENRE. Os atributos da tabela MUSIC são o ID (Chave primária) e a localização da música no sistema de arquivos. Os atributos da tabela MUSIC_INFO são as chaves primárias das tabelas

MUSIC e GENRE, o título da música, o nome do artista, o nome do álbum, o bitrate e o sample rate da música. Os atributos da tabela GENRE são o ID (Chave primária) e o nome do gênero. Com essa descrição textual é possível perceber que a tabela que faz o relacionamento entre as músicas e seus respectivos gêneros é a tabela MUSIC_INFO, dessa forma, se fosse necessário que uma mesma música tivesse mais de um gênero possível, isso poderia ser facilmente adequado no sistema. Outro motivo pelo uso de banco de dados para armazenar as informações é para garantir a integridade dos rótulos associados as músicas cadastradas no sistema.

Além dessas três tabelas principais, o framework possui outras tabelas referentes as características extraídas de diferentes segmentos da música, sendo uma tabela para cada segmento. A vantagem de utilizar tabelas para armazenar as características de diferentes segmentos é que uma vez que foi implementado o método para gerar em algum formato de saída, como por exemplo, arquivos no formato .arff (utilizado pela ferramenta WEKA[10] para mineração de dados), ele pode ser aplicado a qualquer tabela que contenha características extraídas do sinal de áudio.

No que diz respeito a reprodutibilidade dos experimentos, com essa abordagem, todas as músicas disponíveis na base de dados tem disponíveis as informações de Artista e Título. Com essas informações é possível criar junto com os arquivos arffs, gerados para os experimentos, uma lista das músicas utilizadas na mesma ordem em que elas vão ser utilizadas pelo módulo de classificação. O arquivo utilizado para armazenar essa lista é chamado de SAL (Song Artist List). O SAL é uma melhor forma de representar essa informação por três motivos:

1. Algumas vezes artistas diferentes interpretam as mesmas músicas (porém, as vezes, até mesmo em ritmos diferentes), logo utilizar apenas o Título da música não é suficiente;
2. Utilizar o ID da música fornecido pelo sistema não é confiável, pois se por algum motivo for necessário re-cadastrar todas as músicas, elas dificilmente vão ser cadastradas na mesma ordem em que foram cadastradas originalmente;
3. Pode ser que ao observar a lista das músicas utilizadas seja mais fácil de interpretar os resultados obtidos.

Já um módulo para extração das características e seu armazenamento em banco de dados é uma opção interessante não apenas visando a reprodutibilidade dos experimentos, mas também em relação ao tempo computacional necessário para calcular as características de cada música. Atualmente o framework utilizado para extração de características é o Marsyas [8] que extrai um total de 30 características referentes a textura timbral (*Timbral texture*), relacionadas à batida (*beat-related*) e relacionadas às variações da frequência da vibração (*pitch-related*). Características de textura timbral incluem a média e a variância do centróide espectral, do *rolloff* espectral, do fluxo espectral, das taxas de cruzamento zero, dos Coeficientes Cepstrais de Frequência-Mel, e da baixa energia. Características relacionadas à batida incluem as amplitudes relativas e as batidas por minuto. As características relacionadas ao *pitch* incluem os períodos máximos do pico do *pitch* nos histogramas. Estas características formam vetores de trinta dimensões (Textura Timbral: 9 STFT + 10 MFCC; Ritmo: 6; *Pitch*: 5).

2.4. Distribuição

Um requisito importante apresentado no trabalho de McKay et al. [7] é sobre como realizar a distribuição de bases de dados musicais, pois as músicas comerciais possuem leis de direitos autorais que não permitem que elas sejam livremente distribuídas pela internet.

Uma solução apresentada no trabalho de McEnnis et al. [5] é baseada na centralização de diversas bases de dados musicais e seu acesso através de um framework que possibilite aos pesquisadores obterem informações sobre as músicas cadastradas nesse sistema. Esse framework é denominado OMEN (*On demand Metadata ExtractioN*) e sua principal funcionalidade é fornecer uma interface onde o pesquisador seleciona quais as características que deseja trabalhar e a forma como elas devem ser extraídas das músicas disponíveis. Atualmente este framework utiliza o JAudio [6] para extração de características. Caso o pesquisador deseje utilizar um método que não exista na plataforma, existe a opção de submeter os códigos fontes em Java para realizar a extração. Dessa forma, os pesquisadores obtêm as informações que precisam das músicas, através de meta-informações textuais, sem precisar ter acesso aos arquivos de áudio.

A Latin Music Database ainda não está disponível através do OMEN, mas os procedimentos necessários para incluí-la estão sendo tomados. Informações sobre como obter a Latin Music Database, assim como os vetores de características previamente computados das músicas utilizando o Marsyas podem ser encontrados em <http://www.ppgia.pucpr.br/~silla/lmd/>.

3. Conclusões

Neste trabalho foi apresentada a Latin Music Database, uma base de dados inédita e desenvolvida originalmente para a tarefa de classificação automática de gêneros musicais.

A Latin Music Database juntamente com o framework desenvolvido atendem a todas as características desejadas listadas no trabalho de McKay et al. [7], pois a base contém 3.160 músicas comerciais de 10 gêneros musicais no formato MP3. O uso de tabelas em banco de dados para armazenar as informações das músicas permitem que tanto novos campos sejam adicionados facilmente como atribuir mais de um gênero para a mesma música. O processo para realizar a atribuição de gêneros as músicas é inédito, e consiste no uso da percepção humana de como as músicas são dançadas. Para realizar o procedimento de rotulação foram consultados professores de dança de salão com mais de dez anos de experiência. Dessa forma, tornando o conhecimento de gêneros musicais tácito desses especialistas em conhecimento explícito. As características são extraídas utilizando o framework Marsyas e estão disponíveis no formato .arff. Os vetores de características previamente computados das músicas da Latin Music Database estão disponíveis em <http://www.ppgia.pucpr.br/~silla/lmd/>.

Como trabalho futuro, além da inclusão de novos gêneros musicais, a base atual deve ser classificada de acordo com uma hierarquia de gêneros. Por exemplo, o gênero forró seria o gênero principal dos subgêneros: xote, xaxado e baião.

Referências

- [1] J. J. Aucouturier and F. Pachet. Representing musical genre: A state of the art. *Journal of New Music Research*, 32(1):83–93, 2003.
- [2] M. Grimaldi, P. Cunningham, and A. Kokaram. A wavelet packet representation of audio signals for music genre classification using different ensemble and feature selection techniques. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 102–108, Berkeley, California, 2003. ACM Press.
- [3] H. Homburg, I. Mierswa, B. Möller, K. Morik, and M. Wurst. A benchmark dataset for audio classification and clustering. In *Proceedings of the Sixth International*

Conference on Music Information Retrieval (ISMIR'05), pages 528–531, London, UK, 2005.

- [4] T. Li, M. Ogihara, and Q. Li. A comparative study on content-based music genre classification. In *Proceedings of the 26th annual international ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 282–289, Toronto, Canada, 2003.
- [5] D. McEnnis, C. McKay, and I. Fujinaga. Overview of on-demand metadata extraction network (omen). In *Proceedings of the Seventh International Conference on Music Information Retrieval (ISMIR'06)*, 2006.
- [6] D. McEnnis, C. McKay, I. Fujinaga, and P. Depalle. Jaudio: A feature extraction library. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR'05)*, pages 600–603, London, UK, 2005.
- [7] C. McKay, D. McEnnis, and I. Fujinaga. A large publicly accessible database of annotated audio for music research. In *Proceedings of the Seventh International Conference on Music Information Retrieval (ISMIR'06)*, 2006.
- [8] G. Tzanetakis and P. Cook. Marsyas: A framework for audio analysis. *Organized Sound*, 4(3):169–175, 1999.
- [9] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.
- [10] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2nd edition, 2005.