Hybrid Method for Automatic Music Labeling

Irapuru Haruo Flórido
Campus Curitiba
Instituto Federal do Paraná
Curitiba, Brasil
irapuru@ifpr.edu.br

Roberto Tadeu Raittz
dept. Bioinformatic
Universidade Federal do Parana
Curitiba, Brasil
raittz@ufpr.br

***Abstract* --** Automatic music labeling on large-scale bases is a premise to provide systems of music recommendation, important subject in digital world, demanding countless research in music information retrieval field. Although there are large-scale musical bases, such as Million Song Dataset (MSD), that have low-level label signal audio, descendant from audio signal, they are weakly labelled, that is, songs labels may be incomplete at a high level regarding emotion, vocals and instrument. This work aims to present the Music Label Miner (MLM), a hybrid method based on grouping, genetic algorithm and statistical correlation, which generates mappings and possible inferences of high-level labels based on audio signal, through the relationship of a Large-scale base, MSD, with a lower-dimensional Ground Truth reference base. By applying the proposed method, it will be possible to label songs automatically, which contain only low-level labels and, by the models generated from the method, reach high-level labels. The method is composed by: (i) selection and preprocessing of MSD high and low level data, (ii) reference data set called CAL500exp, (iii) MSD data grouping, (iv) CAL500exp vectorization, (v) relationship of vectorized and grouping datasets, (vi) statistical correlation, (vii) mapping, (viii) visualization of selected data characteristics, (ix) generation of models and (x) inference of high and low-level labels.

Index Terms—music labeling, music information retrieval, method hybrid, machine learning

I. INTRODUCÃO

A rotulação automática de músicas é um tema de pesquisa dentro do domínio da Music Information Retrieval (MIR), e busca alcançar certo nível de maturidade de pesquisa e conquistar sua independência dentro da área [2]. Embora o projeto, Music Genome Project [6], seja de 1999, de maneira implícita deu origem a uma série de rotulações de músicas. Mesmo sendo um processo manual com a participação de musicólogos, havia a clareza da suma importância ao fato, demandando estudos de Rotulação automática. A partir da Web 2.0, mais especificamente com advento do comércio eletrônico de músicas digitais, mudou-se a forma como as pessoas faziam a aquisição de suas músicas [11]. Devido a isso, tarefas como busca, recuperação, indexação, extração e sumarização automática dessas informações se tornaram problemas importantes sobre os quais muitas pesquisas têm sido realizadas. Nesse contexto uma área de pesquisa que tem crescido nos últimos anos é a de recuperação de informações musical e associada a ela a Rotulação Automática que visa criar ferramentas capazes de organizar e gerenciar essa grande quantidade de informações e de forma automática [8]. A grande motivação para investigar novos métodos para rotulação automática de músicas é aumentar sobremaneira as possibilidades da rotulação das músicas com informações de novas bases de larga escala, como a *MSD* [9], tendo como referência bases menores, como a *CAL*500*exp* [17], fortemente rotuladas com informações de alto nível e o uso de métodos híbridos de Aprendizagem de Máquina, diferenciados daqueles propostos geralmente pela literatura. Além disso, tem como motivação a redução da atuação manual neste processo, muito custoso em termos de tempo e de recursos humanos especializados. O aspecto original nesta tese está associado à questão da inferência dos rótulos de alto nível para faixas de músicas que se tem somente a informação de baixo nível, sinal acústico, disponíveis nos repositórios de informações musicais de larga escala, como a *MSD*, que serão empregados neste trabalho.

II. MÉTODO HIBRIDO *MLM*

Empregamos no método proposto uma solução híbrida de aprendizagem de máquina (AM) com a combinação de algoritmos não supervisionados de agrupamentos, método de correlacionamento estatístico, algoritmos evolucionários e classificadores supervisionados. O uso desses algoritmos tem a finalidade de gerar modelos para classificar faixas musicais a partir das características extraídas dos rótulos de baixo nível. Posteriormente, será realizada a rotulação automática, determinada pela inferência e seleção dos rótulos de alto nível para as faixas de músicas dentro dos modelos estabelecidos pelo método *MLM*. Antes de avançarmos na explicação do *MLM*, faz-se mister estabelecer a padronização de dois termos relevantes e particulares do método. São os rótulos de alto nível e rótulos de baixo nível, nominados de RAN e RBN, respectivamente. Os RAN estão relacionados com os rótulos de emoção, instrumentos e vocal das músicas e os RBN estão associados ao sinal de áudio e diretamente aos descritores acústicos Beat, Pitch, Timbre e Loudness.

O método *MLM* tem como base três macroprocessos que abordam os seguintes aspectos Fig.1: A) seleção das características (Feature Selection), pré-processamento, normalização das informações e atributos e correlacionamento estatístico; B) redução de dimensão; e C) análise dos mapas e aprendizagem de máquina com a geração de modelo. Serão detalhados as particularidades de cada módulo e os devidos processos que a compõe. A composição principal do *MLM* é baseada na aplicação de um método híbrido com o uso de algoritmos não supervisionados ($K - means$), correlacionamento estatístico de Pearson, Algoritmo genético (AG) e supervisionados (MLP) para geração de modelos. A aplicação destes algoritmos e métodos está inserida nos

processos macros do *MLM* e a forma como eles interagem é explicada como segue.

As fontes de informação abordadas neste método são duas, cujas propriedades são similares e que pelas suas características determinam a estrutura básica do método, a primeira uma Base de Larga Escala (BLE) e de grande dimensão da MSD e a segunda, a CAL500 uma base menor com o objetivo de servir de referência, Ground Thruth (GT) com rótulos de alto e baixo nível bem determinados. Quais dados, de qual base e como serão selecionados e normalizados? São descritos a seguir. Os aspectos musicais que os humanos usam para descrever a música são basicamente pitch, loudness, duração e timbre [5]. Existem várias abordagens para definir os descritores acústicos de áudio como entrada para um algoritmo de AM. No *MLM*, será aplicado, como mencionado anteriormente, o Beat, Pitch, Timbre e Loudness. Desse modo, nos apropriamos dos elementos e da forma de aprendizagem humana e nos deparamos com a aprendizagem de máquina. Em resumo, os atributos acústicos do sinal de áudio, Beats, Pitches, Timbre e Loudness são aplicados no processo de agrupamento, conforme passo descrito no passo "Geração de agrupamentos exploratórios "do método.

A. Seleção de Características

Determinadas as características dos RAN e RBN, que serão empregadas para desenvolver o método, nesta seção, discorremos sobre o pré-processamento dos dados para serem utilizados como entrada no algoritmo de agrupamentos, correlação estatística, algoritmo genético e classificadores supervisionados. Alguns dados necessitam de um pré-processamento antes de serem aplicados aos algoritmos mencionados para eliminação de "sujeira". Os dados podem estar incompletos, com ausência de valores, erros aleatórios, valores aberrantes (outliers), inconsistentes e outras situações não previstas. Como foram adotadas duas bases de dados a MSD e a CAL500 como fonte de informação, Fig.1 o mesmo processo é aplicado aos RBN das duas bases, tendo em vista que a extração dos dados das músicas é muito variável pela particularidade da quantidade de tempo da faixa de música da qual foram extraídas. Esta situação gera vetores com quantidades diferentes de dados extraídos das faixas de músicas e, em decorrência disso, foram estabelecidos métodos de padronização para os RBN. A normalização dos valores dos vetores das características originadas dos RBN seria transformar os valores em uma escala adequada para melhorar o desempenho dos algoritmos de agrupamentos e classificação. Devido à grande variação dos valores obtidos por meio da extração de características, em especial do sinal acústico, os valores foram normalizados pelo limite inferior 0 e o limite superior 1. A quantidade total de elementos dos vetores de cada faixa de música, conforme a Tab. I, anteriormente apresentada é de 63. Resumindo, os RBN de cada faixa de música são representados por um vetor de tamanho 63 da MSD como também da CAL500. Com os dados padronizados e normalizados das duas bases de dados, procede-se a geração de agrupamentos com os dados da base MSD.

Tabela I
GERAÇÃO DE AGRUPAMENTOS DA MSD

K-means	Centroid	# *LLL*
5	5	63
8	8	63
13	13	63
21	21	63
34	34	63
55	55	63
89	89	63
144	144	63
233	233	63
377	377	63
610	610	63

1) Geração de agrupamentos exploratórios da MSD: O agrupamento aplicado a um conjunto de dados, com determinadas características, tem como finalidade reconhecer padrões de algo que é perceptível numericamente. E esta percepção não se dá no âmbito da compreensão humana, mas sim computacionalmente, por meio de um algoritmo de agrupamento, sobretudo quando se tem uma enorme quantidade de amostras, caso particular da MSD. O método *MLM* tem como premissa a geração de agrupamentos necessários para identificar a concentração e convergência das características dos RBN das trilhas de músicas. Como o requisito do *MLM* é trabalhar com uma base larga escala, a geração destes agrupamentos ocorre a partir dos dados a base MSD. Com as características definidas dos RBN da base de larga escala da MSD, geram-se agrupamentos aplicando o algoritmo não supervisionado $K-means$. Os parâmetros de entrada utilizados para o $K-means$ são:

- Características RBN (63);
- *K* (5, 8, 13, 21, 34, 55, 89, 144, 233, 317 e 610);e
- Faixas de músicas úteis da MSD (880mil).

O valor de *K* para os agrupamentos baseia-se na sequência de Fibonacci, presente em várias áreas da natureza e da música também. Daí a motivação adotada para a sequência de números. O valor de *K* na quantidade de 11 gera ao todo 1589 possibilidades de agrupamentos, permitindo, deste modo, uma maior diversidade e distribuição dos dados, importantes para o método no que se refere à associação dos RBN da CAL500 com os centroides dos agrupamentos da MSD. Ver diagrama Fig 1.

2) Associação dos agrupamentos dos centroides com os RBN da CAL500: Os centroides gerados após o processo de agrupamento dos *RBN* da MSD são armazenados para, na continuidade do processo, promover a associação com os RBN da CAL500, conforme está representado na Fig.1. Essa associação é feita por distância Euclidiana de cada uma das músicas da CAL500; isto é, dos RBN com os centroides do RBN de cada agrupamento da MSD. Nesse momento da aplicação do método, vamos alterar a nominação do RBN para RBNa, porque associamos os RBN da CAL500 e da MSD aos

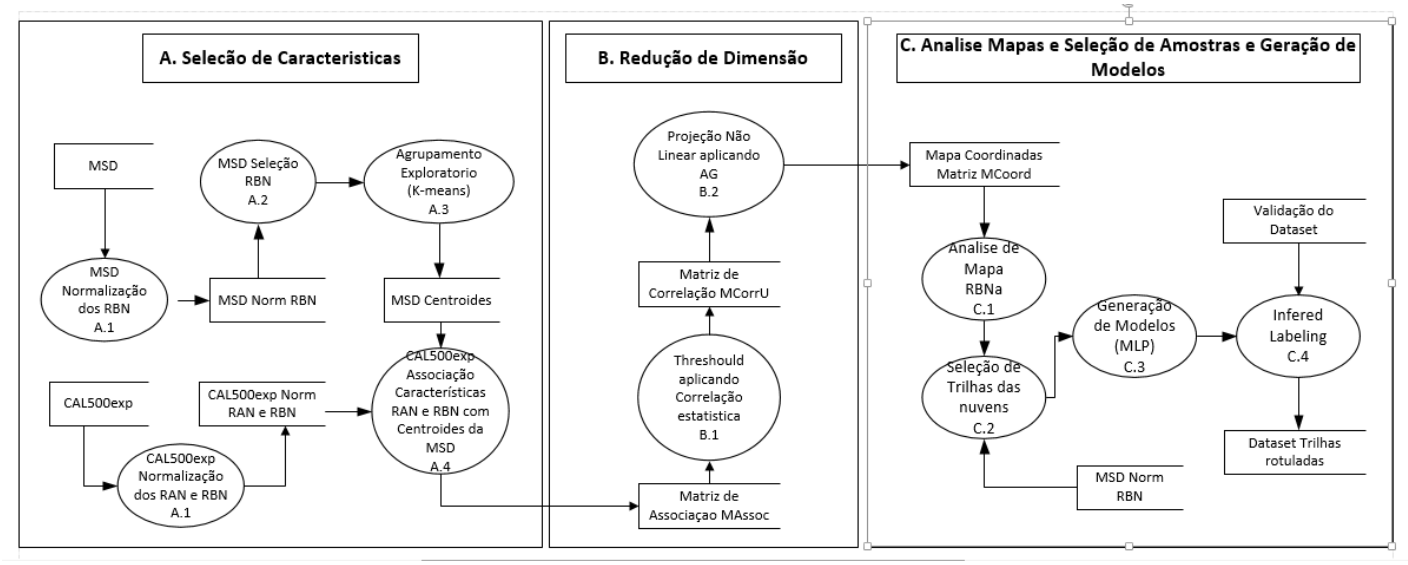

Figura 1. Mapa de Conceitual Geral do Método *MLM*.

centroides dos agrupamentos, ou seja, o "a" representa agrupamentos resultando em Rótulos de Baixo Nível agrupados (RBNa). Um detalhe importante é o produto final do processo que resulta em uma matriz com informações de RAN e RBNa. A relevância se dá porque, nesta matriz de dados, são associados os RAN da CAL500 e os RBN da MSD que foram transformados em centroides e, na sequência, foram associados por distância euclidiana com os RBN da CAL500. Esta associação será uma matriz de $MAssoc_{m,n}$ onde m = nº de trilhas da CAL500 e $n = {}^{P}ki + {}^{P}RAN$.

A matriz resultante $MAssoc_{m,n}$ tem todos os elementos em dois estados, isto é, os elementos são representados por 0 ou 1, tanto os elementos dos RAN como o RBNa. A partir desta etapa, passamos ao cálculo da Matriz de Correlação de Pearson descrito na seção seguinte.

3) Correlação Estatística de Pearson: O emprego do método de correlacionamento estatístico de Pearson, no *MLM* tem alguma restrição porque se emprega a correlação com o objetivo de fazer um threshold, limiarização das informações de um conjunto de dados de referência GT, com os RAN e RBNa da matriz de associação $MAssoc_{m,n}$ calculado no passo anterior. As correlações aceitáveis entre os RAN e RBNa adotados para o método estão no intervalo de 15% a 35% porque neste percentual é possível separar os elementos com o objetivo de filtrar aquelas correlações que estão abaixo do percentual fixado. E os elementos da matriz de correlação são calculados a partir da equação 1. Ressalta-se que a aplicação da formula se dá a cada coluna da matriz $MAssoc_{m,n}$ no intervalo de n = 1,67 para cada elemento de n = 68,1656, assim cada elemento de RAN tem um valor de correlação com o RBNa da matriz. A matriz de Correlação $MCorr_{m,n}$ tem uma dimensão menor que a matriz $MAssoc_{m,n}$, em função do critério do valor percentual mínimo de correlação. O passo seguinte ainda, nesta fase seria achar a Matriz de complemento de 1 da Correlação $MC1Corr_{m,n}$, devido ao fato que na fórmula de Pearson o resultado das correlações varia de -1 a 1. Então, para trabalharmos somente com valores positivos, calculamos a referida matriz $MC1Corr_{m,n}$ que será objeto de entrada para o passo da redução de dimensão.

$$\rho = \frac{cov(X,Y)}{\sqrt{var(X) * var(Y)}} \quad (1)$$

B. Redução de Dimensão

Neste passo do método *MLM*, chegamos a um ponto fundamental, onde será necessário, a partir da matriz $MC1Corr_{m,n}$, gerar o mapa de redução de dimensão com a representação em plano bidimensional, com coordenadas cartesianas das correlações dos RAN e RBNa. Por este, processo podemos visualizar os rótulos bem como visualizar as distâncias entre eles. Isso permite, na sequência, selecionar pelos grupos as respectivas trilhas de músicas contidas em cada agrupamento para gerar os modelos de rotulação automática. Como descrito na visão geral do método, conforme a Fig.1, a aplicação da redução da dimensão serve como um meio para determinar os mapas com os relacionamentos dos RAN e RBNa e o diagrama do processo destaca dois passos para consecução do objetivo, a geração do correlacionamento útil e aplicação do AG com os dados do correlacionamento útil.

1) Aplicação do Algoritmo Genético para a Redução de Dimensão: O uso do AG no contexto do *MLM* seria como a aplicação de uma ferramenta de grande capacidade de geração de valores para determinada função de ajuste. Por esse motivo, foi lançado mão do AG porque a solução algébrica de uma matriz de correlação de dimensão (1656x1656) demandaria um alto custo computacional. Nesse passo, o AG recebe como entrada parâmetros para gerar coordenadas bidimensionais, em função de cada elemento da matriz de Correlação Util $MCorrU_{m,n}$ determinado no passo anterior, ou seja, para cada valor de correlação é retornado do AG uma Matriz de

Coordenadas $MCordU_{m,n}$. Esta matriz permitirá calcular a distância euclidiana entre cada elemento produzindo uma Matriz de Distâncias $MDist_{m,n}$. que permitirá avaliar o custo do AG e verificar a otimização. Os parâmetros iniciais para execução do AG foram os seguintes: a) população inicial: 600; b) elitismo: 20; c) taxa de Mutação: 0,05; d) gerações: n (variável de acordo com o experimento); e) a determinação do tamanho do cromossomo é em função da dimensão da matriz de correlação $MCorrU_{m,n}$ multiplicado por 16, ou seja, o tamanho cromossoma é determinado por: $V\,ectorCrom_{m,n} = 16$ x (m,n); f) após submeter o AG com o cromossomo determinado no item a, transforma-se o vetor resultado em coordenadas par a par (parwise), isto é, gera-se para cada número de elementos da matriz de correlação $MCorrU_{m,n}$ um par de coordenadas; g) função de ajuste (Fitness). A função de ajuste Custo é o cálculo do valor de menor custo resultante da somatória de todos os elementos da matriz de correlação $MCorrU_{m,n}$ subtraindo da somatória de todos os elementos da matriz $MDist_{m,n}$. O processo de ajuste encerra com a determinação do menor custo com número de gerações que foi previamente determinado.

$$Custo = \sum_{m}^{n} M_{corrU} - \sum_{m}^{n} M_{dist} \tag{2}$$

2) Redução de dimensão e geração de mapas: Após encontrar a matriz de coordenadas $MCord_{m,n}$ determinado pelo AG torna-se possível representar um espaço bidimensional de todos os RAN e RBN originados da matriz de correlação $MCorrU_{m,n}$, possibilitando a geração de mapas bidimensionais, como a indicada na Fig.2. O gráfico A da figura apresenta os RAN e RBN com os nomes e o segundo, apenas os pontos, permitindo uma visualização mais exata da proximidade dos rótulos Porém, na Fig.2, com os nomes podemos identificar quais RAN e RBNa estão aglomerados em determinadas nuvens. Na Fig.2 , identificamos a formação das nuvens com os rótulos de RAN e RBNa e da nuvem desta será possível gerar modelos pela análise dos mapas.

C. Análise dos mapas, seleção das amostras e geração dos modelos

Nessa etapa da aplicação do método, estão os passos finais para chegar ao objetivo principal que é rotulação automática propriamente dita das músicas, conforme indicado no diagrama Fig.1, que compreende: análise dos mapas, seleção das amostras e geração dos modelos de rotulação. A geração de mapas de rótulos em um espaço bidimensional permite, pelo critério de distância euclidiana, a visualização de nuvens distintas de agregação de RAN e RBNa, ver Fig.1, de modo que as nuvens estão particularmente separadas por cores diferentes para melhor visualização. As separações das nuvens são encontradas aplicando-se novamente o algoritmo de *K–means* e com os centroides de cada nuvem verificamos os rótulos mais próximos de cada nuvem. As informações dos RAN e RBNa, contidas em cada nuvem, dão subsídios para gerar modelos utilizando classificadores supervisionados para a rotulação

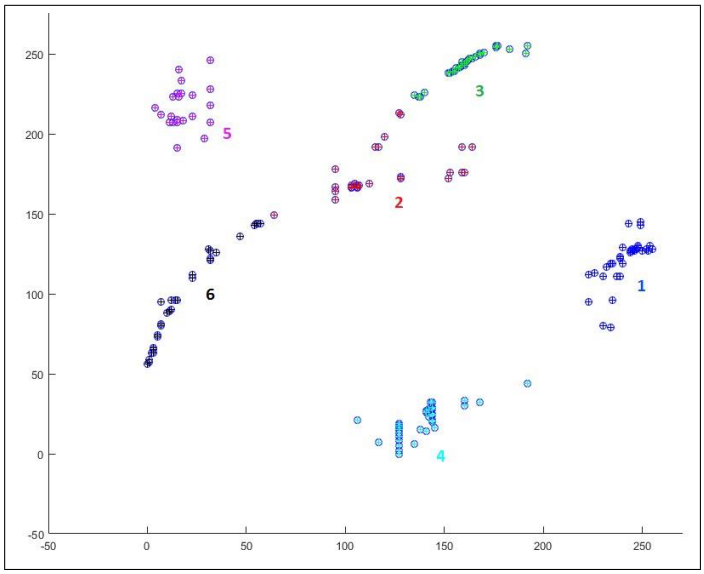

Figura 2. Mapa das Nuvens de Rótulos obtido com a aplicação do AG com 121k gerações com seis nuvens distintas.

automática das músicas. Nesta fase, a seleção dos RBNa presentes em cada nuvem se torna um fator importantíssimo para a geração de modelos, porque as trilhas de músicas relacionadas com estes RBNa serão utilizadas para geração dos dados para criação de modelos por meio de classificadores supervisionados. Na Tab.II temos as distribuições dos RAN e RBNa de cada nuvem, relacionados na . Deste modo, identificamos a separação dos rótulos nas nuvens.

A partir do RBNa, por exemplo, na nuvem 2 o RBNa de número 13/3, destacado, está associado ao agrupamento 13, de maneira que isto indica que as músicas contidas neste e nos outros agrupamentos subsequentes compõem as amostras da

Tabela II
RESULTADOS DA NUVEM 2 COM OS RÓTULOS *RAN* AND *RBNa* EXPERIMENTO DA NUVEM 2

Nuvem	Tipo do Rótulo	Label Id	Nome Rótulo
2	*RAN*	3	EMT – Bizarre
		20	INS – Ambient Sound
		23	INS – Drum Machine
		29	INS – Harmonica
		34	INS – Sample
		36	INS – Sequencer
		38	INS – Synthesizer
		44	VOC – Altered Effects
		53	VOC – Monotone
		55	VOC – Screaming
	LLLa	81	13 / 03
		90	21 / 08
		96	34 / 04
		106	34 / 34
		110	55 / 14
		114	55 / 39
		119	55 / 50
		120	55 / 53

nuvem “2” que possuem RAN características específicas como: emoção (bizarro), instrumento (Som ambiente, bateria, gaita, sample, sequenciador e sintetizador), e vocal (alterado com efeito, monotom e gritado). Assim, todas as músicas contidas nos RBNa, isto é, nos agrupamentos da nuvem 1, em azul na Tab.II têm as características do RAN da nuvem 2. Este processo é repetido para todas as outras nuvens do mapa de rótulos. Concluindo o passo, com a determinação dos RBNa, agrupamentos de cada nuvem, temos as amostras para gerar os modelos para classificar outras músicas somente com os RBN.

1) Aprendizagem de máquina com geração de modelos:

Para a geração de modelos de classificação supervisionado com os dados obtidos na anterior devemos adotar um classificador específico para este fim. Recomendamos, de posse das amostras, fazer um comparativo e verificar o resultado da melhor acurácia média de cada tipo de classificador. Escolher no mínimo três tipos de classificadores supervisionados, um baseado em redes neurais, em árvores e estatísticos. O critério estabelecido no *MLM* para seleção das amostras das trilhas de músicas para gerar os modelos é dependente dos seguintes fatores:

- Estar contido nos agrupamentos dos RBNa de cada nuvem;
- Ser selecionado pela moda estatística de cada nuvem quando a música ocorrer em mais de um RBNa; e
- Ter amostras em todas as nuvens.

Após a seleção das amostras, recuperamos da MSD os RBN originais, ou seja, as 63 características do sinal de áudio com a identificação da nuvem que será a classe para que o classificador MLP possa gerar o modelo.

III. Experimento com a Base de dados de Larga Escala MSD

Os experimentos tiveram como abordagem a aplicação integral do método *MLM* com todas as fases citadas na seção anterior de modo a encontrar os modelos que convergem para rotulação automática das músicas

A. Agrupamento dos rótulos de baixo nível da MSD

Procedemos a Seleção de Características fazendo os agrupamentos dos dados RBN da MSD com os seguintes parâmetros, conforme vemos na Tab.II. Os agrupamentos são gerados de toda a base MSD para cada valor de *K*, ou seja, há um valor do centroide para cada agrupamento aplicando-se o $K-means$. Conforme temos na Tab.II, há 11 formas de agrupamentos, de modo que para cada trilha de música são armazenadas 11 possibilidades de agrupamentos, totalizando 1589 opções de agrupamentos. Cabe reforçar que à associação de cada RBN com o centroide dos agrupamentos denomina-se RBNa, que será referenciado durante a descrição do experimento. O número de trilhas de músicas utilizadas da CAL500 após o pré-processamento resultou em 492, então a $MAssoc_{m,n}$ terá a dimensão de $m = 492$ e $n = 67+1589 = 1656$.

B. Associação do RBN da CAL500 com os Centroides

Neste passo do experimento, fazemos a associação dos RBN de todas as trilhas de músicas da CAL500 com os centroides da MSD, para gerar conforme o método, a matriz de Associação $MAssoc_{m,n}$.

Tabela III
Matriz de Associação dos Rótulos RAN e RBNa

Musica 2 Tracks	Sentimento RAN							Agrupamento RBNa				
	Angry	Arousing	Bizarre	Calming	Carefree	Cheerful	Emotional	1	2	3	4	5
A1203	0	1	0	0	1	1	0	1	0	0	0	0
N1201	0	0	0	1	0	0	1	0	0	0	0	1
G1207	0	1	0	0	1	0	0	0	0	1	0	0
Z1220	0	0	1	1	1	0	0	0	0	0	1	0

O valor de *n* é a soma do número de RAN com RBNa. Vale observar que os RBNa estão marcados com ”0 ou ı1 posicionalmente, mas para cada agrupamento só há uma opção marcada, ver Tab.III.

C. Correlacionamento estatístico de Pearson

Com a $MAssoc_{m,n}$, fazemos o cálculo de correlacionamento estatístico de Pearson de cada RAN com cada RBNa e, ao final do processamento, eliminamos as correlações que estão abaixo do percentual mínimo de correlacionamento e no mesmo processamento fazemos também o complemento de um dos valores dos correlacionamentos gerando-se a $MCorrU_{m,n}$. Na Fig.3, é sensível a diferença dos RBNa para um percentual mínimo de correlação para um maior. Quando a correlação de Pearson aumenta, os dados possuem uma correlação mais alta. Para determinarmos o protocolo dos experimentos, adotamos o valor de mínimo de correlação de 20% equilibrando-se desta forma a quantidade de RBNa, porque um valor de RBNa alto reflete em uma quantidade de elementos matriciais. E isso demanda um alto custo computacional de processamento principalmente na redução de dimensão onde usa-se o AG, conforme vemos a seguir.

D. Redução de Dimensão

Conforme a seção 2.8 do método, o caminho encontrado para representar as correlações dos rótulos RNA e RBNa num plano bidimensional é por meio da aplicação do AG, ferramenta que transforma estas correlações em coordenadas. Para tanto, utilizamos a matriz $MCorrU_{m,n}$. Como adotamos o valor de ρ_{min} = 20%, a matriz tem a dimensão de $MCorrU_{436,436}$. Além de representar as correlações em espaço bidimensional, o método neste passo visa a determinar a convergência dos rótulos em aglomerações que denominamos de nuvem. Experimentalmente, quanto maior o número de gerações aplicado ao AG, maior é a convergência em nuvens dos RAN e RBNa.

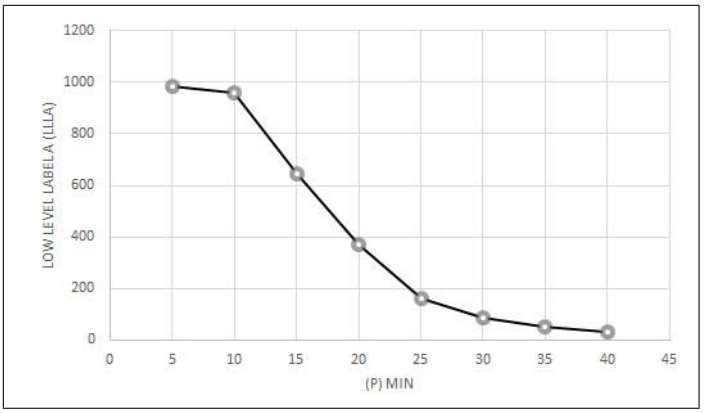

Figura 3. Percentual de Correlação (ρ x *RBNa*)

Percebemos claramente a convergência em nuvens distintas, de acordo com a Fig.2, a cada geração diferente do AG, isto resulta das características comuns dos rótulos derivados do sinal de áudio das músicas ou RNB. Apresentadas as considerações primárias do experimento K5610c121K, fazemos a abordagem detalhada dos resultados obtidos. Inicialmente, a partir dos resultados da redução da dimensão, que gerou o mapa apresentado na Fig.2, selecionamos os RBNa de cada nuvem. Por exemplo, ver Tab.II, da nuvem 2 são selecionados 43 RBNa que são utilizados para compor o conjunto de amostras para gerar os modelos do classificador supervisionado MLP. Relembrando, estes 43 RBNa são os agrupamentos gerados no passo descrito na seção 5.2 e com estes RBNa voltamos a MSD e selecionamos todas as faixas de músicas que estão contidas neste RBNa. Por exemplo, ainda na nuvem 2, o primeiro RBNa é o "j_13_ids_3", ver Tab.II, ou seja, buscamos todas as faixas de músicas na MSD que estão no grupo 13/3 no total 71392 faixas de músicas. Esse procedimento se repete e aplicamos a todos os RBNa da nuvem, totalizando desta forma os quantitativos listados na Tab. IV Importante salientar que a diferença de quantitativo inicial para o final é decorrente da aplicação da Moda Estatística, isto é, como uma faixa de música pode estar em vários RBNa, consideramos no experimento somente as faixas de músicas para a nuvem que esteja dentro da Moda Estatística.

E. Analise dos Resultados

Para escolha do melhor classificador supervisionado para gerar o modelo de rotulação automática, experimentamos três opções de classificadores: o MLP, baseado em redes neurais, o J48 baseado em árvores e o Naive Bayes baseado em estatísticas. Em função dos resultados da melhor acurácia média, ver Tab. V, optamos pelo classificador MLP dado o maior resultado apresentado com o número de geração 121393. Com o conjunto de amostras definidos e a escolha do classificador MLP, podemos verificar ainda na Tab.IV na última linha, que a acurácia do MLP com 121393 resultou em 97,1% sendo a melhor acurácia dentre os experimentos efetuados.

Tabela IV
COMPARAÇÃO DOS RESULTADOS DOS CLASSIFICADORES MLP, J48 E NAIVE BAYES

Configuração dos Experimentos		Classificadores		
GA Gerações	Numeros de Amostras	MLP (%)	J48 (%)	Naïve Bayes (%)
10K	88858	94.4	90.9	94.4
17K	108767	94.5	91.4	94.5
28K	95995	94.8	86.8	94.8
46K	74942	94.6	89.7	94.6
75K	104974	95.1	89.7	95.1
121K	86122	97.1	88.5	95.1

Neste comparativo podemos afirmar que quanto maior o número de gerações aplicados ao AG a acuracia do modelo MLP é maior. A validação do modelo gerado pelo classificador MLP é submetido pelo conjunto de amostras da Last FM, que são amostras rotuladas por usuários do sítio de mesmo nome. Os passos realizados para validação resultaram na Tab.V de valores, cuja acurácia é calculada pelo quociente da presença do rótulo da faixa de música na nuvem pelo total do conjunto de amostras.

Analisando os resultados da validação, concluímos que no geral o percentual de acerto é de 28,8%, quociente da quantidade total de amostras pela quantidade total de acertos, (106399 / 30598), ver Tab.V. Embora o valor do acerto varia com o número de amostradas contida em cada rótulo. O problema que encontramos na validação é que os rótulos nas amostras de validação não estão proporcionalmente distribuídos com os tipos de rótulos de emoção, instrumento e vocal. Porque o conjunto de dados Cal500 base GT [17] foi referência para a criação do modelo de aprendizagem.

Tabela V
RESULTADOS DA VALIDAÇÃO COM A BASE LAST.FM

Número de Rótulos	Número de Amostras	Acurácia (%)
1	55,441	39.2
2	26,502	19.8
3	11,633	16.8
4	5,794	14.2
5	3,225	14.8
6	1,766	14.3
7	927	8.8
8	499	6.6
9	271	8.1
10	166	1.8

IV. CONCLUSÃO

O método de composição híbrida determinante se mostra fiel ao objetivo principal proposto, ou seja, a rotulação automática de faixas de músicas em larga escala, com o vínculo das informações de baixo nível provenientes dos sinais de áudio com as informações de alto nível advindos da emoção, do instrumento e do vocal. As informações de alto nível dos

trechos de músicas, como emoções, por exemplo, têm características de alta subjetividade ao serem tratadas e inferidas no método, causando algumas distorções nos resultados. A base de referência Ground-Truth, utilizada no método CAL500 [17], é, de certa forma, limitada frente à quantidade e a diversidade das músicas contidas na base de larga escala MSD. No momento em que houver bases de músicas rotuladas similar ao projeto Pandora [6] poderemos ter uma base real de referência. Por questão de direitos autorais, os RBN dos trechos de músicas limitam de algum modo a recuperação das informações pela limitação do uso de trechos de músicas. Devido à grande quantidade de amostras do conjunto de dados de larga escala nas diferentes fases de pré-processamento e processamento dos experimentos, devemos considerar o elevado custo computacional do método. A contribuição deste método visa também a abranger outras áreas de pesquisa, além da recuperação de informação de música, pela característica estrutural do *MLM*.

REFERÊNCIAS

[1] Cook, Perry R.,Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics, 1999 ,0-262-03256-2, MIT Press Cambridge, MA, USA.

[2] Downie, J S and Cunningham, S J,Toward a theory of music information retrieval queries: System design implications,Proceedings of the 3o. International Society for Music Information Retrieval Conference (ISMIR'02), pp. 13–17,2002.

[3] Ellis, K. and Coviello, E. and Chan, A. and Lanckriet G,IEEE Transactions on Audio, Speech and Language Processing,vol 21, pp. 2554– 2569,A bag of systems representation for music auto-tagging, 2013.

[4] Freitas, Alex A.",Maimon, Oded and Rokach, Lior", "A Review of evolutionary Algorithms for Data Mining","Soft Computing for Knowledge
Discovery and Data Mining",2008,Springer US,Boston, MA,pp 79–111.

[5] Anil K. Jain,Data clustering: 50 years beyond K-means,Pattern Recognition Letters,vol. 31,pp 651-666,2010,Award winning papers from the 19th International Conference on Pattern Recognition (ICPR),2010.

[6] John J. Joyce,Scientific Computing,vol.23,pp. 14–41,Pandora and the Music Genome Project,2006.

[7] Knees, Peter and Schedl, Markus,Music Retrieval and Recommendation: A Tutorial Overview,Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval,2015, Santiago, Chile,pp. 1133–1136,New York, NY, USA.

[8] Lidy, Thomas and Jr, Carlos N Silla and Cornelis, Olmo and Gouyon, Fabien and Rauber, Andreas and Kaestner, Celso A A and Koerich, Alessandro L,On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-Western and ethnic music collections, Signal Processing,vol. 90,2010

[9] McFee, Brian and Bertin-Mahieux, Thierry and Ellis, Daniel P.W. and Lanckriet, Gert R.G.,The Million Song Dataset Challenge,Proceedings of the 21st International Conference on World Wide Web,2012,pp.909–916,New York, NY, USA.

[10] Pachet, François and Zils, Aymeric,Signal Processing, pp. 103–103,Automatic extraction of music descriptors from acoustic signals, vol.10, 2004.

[11] Pan, Jeff Z. and Taylor, Stuart and Thomas, Edward, Reducing Ambiguity in Tagging Systems with Folksonomy Search Expansion,Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications, Heraklion, Crete, Greece,Springer-Verlag,2009.

[12] Shen, Jialie and Meng, Wang and Yan, Shuichang and Pang, HweeHwa and Hua, Xiansheng,Effective Music Tagging Through Advanced Statistical Modeling,Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval,pp. 635–642,New York, NY, USA,2010.

[13] Thierry Bertin-mahieux and Daniel P. W. Ellis and Brian Whitman and Paul Lamere,The million song dataset,In Proceedings of the 12th International Conference on Music Information Retrieval ISMIR,2011.

[14] Turnbull, Douglas and Barrington, Luke and Torres, David and Lanckriet, Gert, Towards Musical Query-by-semantic-description Using the CAL500 Data Set, Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,Amsterdam, The Netherlands, 2007.

[15] Turnbull, D. and Barrington, L. and Torres, D. and Lanckriet, G.,Semantic Annotation and Retrieval of Music and Sound Effects,Trans. Audio, Speech and Lang. Proc.,vol.16,IEEE Press,Piscataway, NJ, USA,2008.

[16] G. Tzanetakis and P. Cook,IEEE Transactions on Speech and Audio Processing,Musical genre classification of audio signals,vol.10,,2002.

[17] Wang, Shuo Yang and Wang, Ju Chiang and Yang, Yi Hsuan and Wang, Hsin Min,Towards time-varying music auto-tagging based on CAL500 expansion, Proceedings - IEEE International Conference on Multimedia and Expo,IEEE Computer Society,vol. 2014-Septe,2014.

[18] A. Theocharis and M. Pierce and G. Tzanetakis, An Empirical Investigation of Stacking for Music Tag Annotation,10th International Conference on Machine Learning and Applications and Workshops,vol.1,2011.