

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS CORNÉLIO PROCÓPIO
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

ADRIANO ALVES DE LIMA

**CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS EM
BASES DE LETRAS DE MÚSICAS EM DIFERENTES IDIOMAS
UTILIZANDO MÉTODOS DE COMBINAÇÃO DE
CLASSIFICADORES**

DISSERTAÇÃO DE MESTRADO

CORNÉLIO PROCÓPIO

2016

ADRIANO ALVES DE LIMA

**CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS EM
BASES DE LETRAS DE MÚSICAS EM DIFERENTES IDIOMAS
UTILIZANDO MÉTODOS DE COMBINAÇÃO DE
CLASSIFICADORES**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Tecnológica Federal do Paraná – UTFPR como requisito parcial para a obtenção do título de “Mestre em Informática”.

Orientador: Prof. Dr. Carlos N. Silla Jr.

CORNÉLIO PROCÓPIO

2016

Dados Internacionais de Catalogação na Publicação

- L732 Lima, Adriano Alves de
Classificação automática de gêneros musicais em bases de letras de músicas em diferentes idiomas utilizando métodos de combinação de classificadores / Adriano Alves de Lima. – 2016.
73 f. : il. ; 31 cm
- Orientador: Carlos Nascimento Silla Junior.
Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Informática. Cornélio Procópio, 2016.
Bibliografia: p. 70-73.
1. Classificação. 2. Identidade de gênero na música. 3. Idiomas. 4. Linguística – Processamento de dados. 5. Informática – Dissertações. I. Silla Junior, Carlos Nascimento, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Informática. III. Título.

CDD (22. ed.) 004



Título da Dissertação Nº 15:

“Classificação Automática de Gêneros Musicais em Bases de Letras de Músicas em Diferentes Idiomas Utilizando Métodos de Combinação de Classificadores”.

por

Adriano Alves de Lima

Orientador: **Prof. Dr. Carlos Nascimento Silla Junior**

Esta dissertação foi apresentada como requisito parcial à obtenção do grau de MESTRE EM INFORMÁTICA – Área de Concentração: Computação Aplicada, pelo Programa de Pós-Graduação em Informática – PPGI – da Universidade Tecnológica Federal do Paraná – UTFPR – Câmpus Cornélio Procópio, às 08h e 30min do dia 17 de dezembro de 2015. O trabalho foi aprovado pela Banca Examinadora, composta pelos professores:

Prof. Dr. Carlos Nascimento Silla Junior
(Presidente – UTFPR-CP)

Prof. Dr. Danlio Sipoli Sanches
(UTFPR-CP)

Prof. Dr. Emerson Cabrera Paraíso
(PUC-PR)

Visto da coordenação:

Carlos Nascimento Silla Junior
Coordenador do Programa de Pós-Graduação em Informática
UTFPR Câmpus Cornélio Procópio

A Folha de Aprovação assinada encontra-se na Coordenação do Programa.

AGRADECIMENTOS

Primeiramente, agradeço a Deus pela benção de chegar a esta fase da vida e estar preparado para muitas outras.

Aos meus pais Wiliam de Alves de Lima e Renilde Ferreira de Lima. Minha amada mãe por seu apoio incondicional e incentivo. E a memória de meu querido pai, que enquanto presente sempre me apoiou e me orientou em minhas decisões.

A minha namorada e parceira de todas as horas, por seu amor, compreensão e ajuda em diversos momentos neste período.

Aos meus companheiros de trabalho de Maringá, Londrina e também do Laboratório de Computação e Tecnologia Musical pela companhia, compartilhamento de conhecimento e momentos divertidos.

Agradeço em especial ao meu orientador Carlos Nascimento Silla Junior pela amizade, orientações no desenvolvimento do trabalho e parcerias incontáveis.

A banca examinadora pela atenção dada a esse trabalho e aos demais professores e servidores da Universidade Tecnológica Federal do Paraná, campus Cornélio Procópio e Londrina, pelo compartilhamento de conhecimento e ajuda nesta jornada.

Agradeço também aos amigos de turma pela parceria nos trabalhos e também pelos momentos de conversa e troca de experiência.

Por fim, peço desculpas aos demais amigos e pessoas não citados que me auxiliaram nesta longa caminhada, meus sinceros agradecimentos.

“Sábio é o ser humano que tem coragem de ir diante do espelho da sua alma para reconhecer seus erros e fracassos e utilizá-los para plantar as mais belas sementes no terreno de sua inteligência.” Augusto Cury

RESUMO

de LIMA, Adriano Alves. CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS EM BASES DE LETRAS DE MÚSICAS EM DIFERENTES IDIOMAS UTILIZANDO MÉTODOS DE COMBINAÇÃO DE CLASSIFICADORES. 74 f. Dissertação de Mestrado – Programa de Pós-graduação em Informática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2016.

A classificação automática de gêneros musicais é uma tarefa importante na área de recuperação da informação musical. Para auxiliar esta tarefa diversos recursos linguísticos de processamento de linguagem natural vem sendo utilizados em letras de músicas. Isto é importante considerando que a classificação baseada em áudio pode ser complementada com características das letras de músicas para auxiliar na melhoria dos resultados. Este trabalho busca realizar a análise de tais ferramentas de processamento de linguagem natural aplicadas as letras de músicas em diferentes idiomas para a classificação de gêneros musicais. Além disso, foram estudadas outras técnicas para melhorar os resultados da classificação, como por exemplo, combinação de características por meio de *early fusion* e *late fusion*.

Para realizar os experimentos foram criadas três bases de dados com letras de músicas em diversos gêneros nos idiomas Latinos (Português e Espanhol), Nórdicos (Dinamarquês, Norueguês e Sueco) e Inglês. Os resultados obtidos demonstram que a combinação das abordagens por meio das técnicas de *late fusion* através do produto da probabilidade e da técnica de *early fusion* proporcionaram os melhores resultados na classificação de gêneros musicais utilizando letras de músicas em diferentes idiomas.

Palavras-chave: classificação, gêneros musicais, diferentes idiomas, letras de músicas, *early fusion*, *late fusion*

ABSTRACT

de LIMA, Adriano Alves. MUSIC GENRE AUTOMATIC CLASSIFICATION IN SONG LYRICS DATABASE IN DIFFERENT LANGUAGES USING ENSEMBLES CLASSIFIERS METHODS. 74 f. Dissertação de Mestrado – Programa de Pós-graduação em Informática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2016.

The automatic music genre classification is an important area in music information retrieval. For assist this task many languages resources of natural language processing has been used in song lyrics. This is important considering that the classification audio-based might be complemented with lyrics features to improve the results. This work analyzes such natural language processing tools employed to song lyrics in different languages for the music genre classification. Furthermore, we studied other techniques to improve the classification results, for example, features combination by mean of early and late fusion methods.

To run the experiments were created three lyrics database in various music genres into the Latin languages (Portuguese and Spanish), Nordic (Danish, Norwegian and Swedish) and English. The results obtained shown that the combination approaches by mean of late fusion with product of the probability and early fusion outperformed the other approaches in music genre classification using song lyrics in different languages.

Keywords: classification, music genres, different languages, song lyrics, early fusion, late fusion

LISTA DE FIGURAS

FIGURA 1	– Esquema de combinação de características por <i>early fusion</i>	23
FIGURA 2	– Esquema de combinação de características por <i>late fusion</i> usando probabilidade máxima	24
FIGURA 3	– Esquema de combinação de características por <i>late fusion</i> usando soma das probabilidades	25
FIGURA 4	– Esquema de combinação de características por <i>late fusion</i> usando produto das probabilidades	26
FIGURA 5	– Nuvem de palavras para o gênero dance da NLD	31
FIGURA 6	– Nuvem de palavras para o gênero pop rock da NLD	31
FIGURA 7	– Nuvem de palavras para o gênero rap da NLD	31
FIGURA 8	– Nuvem de palavras para o gênero pop da NLD	32
FIGURA 9	– Nuvem de palavras para o gênero rock da NLD	32
FIGURA 10	– Nuvem de palavras para o gênero axé da LMLD	33
FIGURA 11	– Nuvem de palavras para o gênero bachata da LMLD	33
FIGURA 12	– Nuvem de palavras para o gênero bolero da LMLD	33
FIGURA 13	– Nuvem de palavras para o gênero pagode da LMLD	34
FIGURA 14	– Nuvem de palavras para o gênero forró da LMLD	34
FIGURA 15	– Nuvem de palavras para o gênero gaúcha da LMLD	34
FIGURA 16	– Nuvem de palavras para o gênero merengue da LMLD	35
FIGURA 17	– Nuvem de palavras para o gênero salsa da LMLD	35
FIGURA 18	– Nuvem de palavras para o gênero sertaneja da LMLD	35
FIGURA 19	– Nuvem de palavras para o gênero tango da LMLD	36
FIGURA 20	– Nuvem de palavras para o gênero blues da LGTZAN	37
FIGURA 21	– Nuvem de palavras para o gênero country da LGTZAN	37
FIGURA 22	– Nuvem de palavras para o gênero disco da LGTZAN	37
FIGURA 23	– Nuvem de palavras para o gênero hip hop da LGTZAN	38
FIGURA 24	– Nuvem de palavras para o gênero metal da LGTZAN	38
FIGURA 25	– Nuvem de palavras para o gênero pop da LGTZAN	38
FIGURA 26	– Nuvem de palavras para o gênero rock da LGTZAN	39
FIGURA 27	– Nuvem de palavras para o gênero reggae da LGTZAN	39
FIGURA 28	– Visão geral das etapas de processamento de letras de músicas	41

LISTA DE TABELAS

TABELA 1	– Características estilísticas	20
TABELA 2	– Resumo dos trabalhos relacionados	29
TABELA 3	– Número de atributos para cada representação sem e com remoção de <i>stopwords</i> para cada base de dados de letras de músicas.	40
TABELA 4	– Mapeamento de caracteres especiais para os idiomas Nórdicos	41
TABELA 5	– Resultados para cada base de dados individualmente	43
TABELA 6	– Resultados para os métodos <i>early</i> e <i>late fusion</i> na LGTZAN	46
TABELA 7	– Resultados para os métodos <i>early</i> e <i>late fusion</i> na LMLD	51
TABELA 8	– Resultados para os métodos <i>early</i> e <i>late fusion</i> na NLD	55
TABELA 9	– Resultados para os métodos <i>early</i> e <i>late fusion</i> na JLD	59
TABELA 10	– <i>Ranking</i> médio dos métodos utilizados	66

LISTA DE SIGLAS

BOW	<i>Bag Of Words</i>
ELF	<i>Ethnic Lyrics Fetcher</i>
JLD	<i>Joint Lyrics Database</i>
LGTZAN	<i>Lyrics GTZAN</i>
LMD	<i>Latin Music Database</i>
LMLD	<i>Latin Music Lyrics Database</i>
MIR	<i>Music Information Retrieval</i>
MIREX	<i>Music Information REtrieval eXchange</i>
NLD	<i>Nordic Lyrics Database</i>
PLN	<i>Processamento de Linguagem Natural</i>
POS	<i>Part-Of-Speech</i>
SMO	<i>Sequential Minimal Optimization</i>
SSD	<i>Statistical Spectrum Descriptors</i>
SVM	<i>Support Vector Machines</i>
SWR	<i>Stopwords Removal</i>
TF-IDF	<i>Term Frequency Inverse Document Frequency</i>

LISTA DE SÍMBOLOS

TF	Frequência do termo em um documento
nD	Número de documentos utilizados
nDF	Número de documentos onde ocorre a representação do termo
Vp	Verdadeiro positivo
Fp	Falso positivo
Fn	Falso negativo
Z	Instância para ser classificada
m	Quantidade de classes possíveis
R	Modelos de representações
x_i	i -ésimo classificador x
q_α	Valores críticos

SUMÁRIO

1 INTRODUÇÃO	12
1.1 CONTEXTUALIZAÇÃO	12
1.2 MOTIVAÇÃO E JUSTIFICATIVA	12
1.3 HIPÓTESE	13
1.4 OBJETIVOS	13
1.4.1 Objetivos Específicos	14
1.5 ORGANIZAÇÃO DO DOCUMENTO	14
2 PROBLEMA	15
3 REFERENCIAL TEÓRICO	17
3.1 MINERAÇÃO DE TEXTOS	17
3.1.1 Remoção de <i>stopwords</i>	17
3.1.2 <i>Stemming</i>	18
3.1.3 Caracteres n-gramas	18
3.1.4 TF-IDF	18
3.1.5 Características estilísticas de texto	19
3.2 APRENDIZADO SUPERVISIONADO	19
3.2.1 Classificador SVM	20
3.2.2 Métodos para a avaliação de modelos	21
3.2.2.1 Validação cruzada	21
3.2.2.2 Métricas para avaliação de resultados	22
3.3 COMBINAÇÃO DE CLASSIFICADORES	22
3.3.1 Early Fusion	22
3.3.2 Late Fusion	23
3.3.2.1 Probabilidade Máxima	24
3.3.2.2 Soma das Probabilidades	25
3.3.2.3 Produto das Probabilidades	25
3.4 TRABALHOS RELACIONADOS	26
4 METODOLOGIA	30
4.1 CRIAÇÃO DAS BASES DE DADOS	30
4.1.1 Nordic Lyrics Database	30
4.1.2 Latin Music Lyrics Database	32
4.1.3 Lyrics GTZAN Database	36
4.1.4 Joint Lyrics Database	39
4.1.5 Detalhes das bases de dados	39
4.2 PROCESSAMENTO DE LETRAS DE MÚSICAS	40
4.2.1 Detecção de idiomas	41
4.2.2 Pré-processamento	41
4.2.3 <i>Stoplists</i> utilizadas	42
4.2.4 Características estilísticas de texto utilizadas	42

5	RESULTADOS SEM COMBINAÇÃO (NO FUSION) PARA OS EXPERIMENTOS COMPUTACIONAIS	43
5.1	RESULTADOS PARA ABORDAGENS SEM COMBINAÇÃO (NO FUSION)	43
5.1.1	Qual representação usar?	44
5.1.2	Qual é o impacto das <i>stopwords</i> ?	44
5.2	AVALIAÇÃO E DISCUSSÃO GLOBAL DOS RESULTADOS	45
6	EXPERIMENTOS COMPUTACIONAIS	46
6.1	RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA LGTZAN	46
6.1.1	Combinando N-Gramas com Representações <i>Stemming</i>	48
6.1.2	Combinando N-Gramas com Características Estilísticas	49
6.1.3	Combinando N-Gramas, <i>Stemming</i> e Características Estilísticas	49
6.1.4	Outras Combinações	50
6.2	RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA LMLD	51
6.2.1	Combinando Representações N-Gramas com <i>Stemming</i>	52
6.2.2	Combinando N-Gramas com Características Estilísticas	53
6.2.3	Combinando N-Gramas, <i>Stemming</i> e Características Estilísticas	54
6.2.4	Outras Combinações	54
6.3	RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA NLD	55
6.3.1	Combinando as Representações N-Gramas com <i>Stemming</i>	57
6.3.2	Combinando N-Gramas com Características Estilísticas	57
6.3.3	Combinando N-Gramas, <i>Stemming</i> e Características Estilísticas	58
6.3.4	Outras Combinações	58
6.4	RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA JLD	59
6.4.1	Combinando N-Gramas com Representações <i>Stemming</i>	61
6.4.2	Combinando N-Gramas com Características Estilísticas	62
6.4.3	Combinando N-Gramas, <i>Stemming</i> e Características Estilísticas	62
6.4.4	Outras Combinações	63
6.5	AVALIAÇÃO E DISCUSSÃO GLOBAL DOS RESULTADOS	63
6.5.1	Avaliação dos Resultados	64
6.5.2	Teste Friedman e Nemenyi	64
6.5.3	Análise dos Testes Estatísticos	65
7	CONCLUSÕES	67
	REFERÊNCIAS	70

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

O estudo das tecnologias musicais tem sido um campo importante em muitas áreas de pesquisa, como por exemplo, em *Music Information Retrieval* (MIR), na qual as pesquisas comumente são direcionadas a recuperação por similaridade musical, classificação de gêneros musicais ou análise de músicas e representação do conhecimento (ORIO, 2006). Em diversos trabalhos (NEUMAYER; RAUBER, 2007; MAYER et al., 2008a, 2008b; MAYER; NEUMAYER, 2009; MAYER; RAUBER, 2010, 2011) foi demonstrado que as letras de músicas são relevantes para auxiliar ou substituir outras abordagens na tarefa de classificação de gêneros musicais. Além disso, as letras de músicas também foram utilizadas para melhorar o reconhecimento de emoção em músicas (HU et al., 2009, 2009; HU; DOWNIE, 2010a, 2010b). Levando em consideração que a música está presente em diferentes países e culturas, naturalmente as letras das músicas são escritas em diferentes idiomas. Entretanto a maioria dos trabalhos desenvolvidos utilizam apenas letras de músicas em inglês. Mais ainda, de acordo com (SILLA JR. et al., 2010; ZANEN; KANTERS, 2010) o uso de letras de músicas geralmente depende de ferramentas de Processamento de Linguagem Natural (PLN) que muitas vezes não estão disponíveis para músicas de outros idiomas, como por exemplo, *Part-Of-Speech* (POS).

1.2 MOTIVAÇÃO E JUSTIFICATIVA

A análise e desenvolvimento de técnicas para a classificação automática de gêneros musicais utilizando letras de músicas em diferentes idiomas são necessários devido ao aumento exponencial de dados musicais nos últimos anos (ORIO, 2006; NEUMAYER; RAUBER, 2007). Além disso, em Bibliotecas de Músicas Digitais é comum usar gêneros musicais para organizar as coleções de músicas. Para os usuários finais dos sistemas a navegação e busca através de gêneros musicais é mais significativa do que recomendações ou similaridade por artis-

tas ou músicas de acordo com (LEE; DOWNIE, 2004). Mais ainda, a classificação automática de gêneros musicais é também importante para aplicações comerciais, tais como: Super Player e Pandora (MCKAY; FUJINAGA, 2006; ORIO, 2006). Isto se deve ao fato do gênero musical ser uma categoria e/ou conceito popular para a classificação das músicas, independentemente da razão, propósito ou critério (FABBRI, 1999a).

O estado da arte não contempla estudos voltados para a classificação de gêneros musicais utilizando letras de músicas em diferentes idiomas, sendo, portanto, focados em letras no idioma Inglês (MAYER et al., 2008a, 2008b; MCKAY et al., 2010; MAYER; RAUBER, 2011; FELL; SPORLEDER, 2014). Esse trabalho abrange a análise de técnicas de classificação utilizando recursos dependentes e independentes de idiomas aplicadas as letras de músicas em diferentes idiomas, além das combinações destas técnicas visando a melhoria da performance de classificação (SILLA JR.; FREITAS, 2009, 2011). As combinações foram feitas por meio das técnicas *early* e *late fusion*. A combinação por *early fusion* realiza a junção das características unimodais antes da etapa de aprendizado de máquina. Em contra partida, os métodos baseados em *late fusion* realizam a combinação de características ou saídas dos classificadores posteriormente a etapa de aprendizado de máquina (SNOEK et al., 2005). Para avaliar as técnicas houve a necessidade de desenvolver três bases de letras de músicas em diferentes gêneros e idiomas, além da junção destas para verificar se os resultados obtidos nas bases individualmente podem ser escalados para bases de dados com maior complexidade através da variação de idiomas e gêneros musicais.

1.3 HIPÓTESE

A hipótese que conduzirá o desenvolvimento deste trabalho é a que se segue: O uso de técnicas de combinação de classificadores leva a melhores taxas no reconhecimento de gêneros utilizando letras de músicas.

1.4 OBJETIVOS

Esse trabalho tem como objetivo principal a análise de técnicas de PLN para a classificação de gêneros musicais utilizando letras de músicas em diferentes idiomas e a avaliação destas técnicas combinadas.

1.4.1 OBJETIVOS ESPECÍFICOS

- Analisar as técnicas de combinação de *early fusion* e *late fusion*.
- Desenvolver novas bases de letras de músicas em diferentes idiomas e gêneros.
- Avaliar os métodos de combinação.

1.5 ORGANIZAÇÃO DO DOCUMENTO

Esse trabalho está dividido em 7 capítulos. Além do capítulo inicial de introdução do trabalho que esclarece a contextualização, motivação e justificativa, objetivos e organização do documento, os demais capítulos são os seguintes:

- **Problema:** neste capítulo é apresentado de forma detalhada qual o problema central desse trabalho. Este capítulo se faz necessário para apresentar o estado da arte em classificação de gêneros musicais utilizando letras de músicas e também para esclarecer de forma detalhada os termos importantes para compreensão do trabalho.
- **Referencial Teórico:** neste capítulo são apresentados os métodos que serão utilizados para realização do trabalho.
- **Metodologia:** neste capítulo são apresentados os protocolos experimentais utilizados. Sendo eles: criação das bases de dados, processamento de letras de músicas e interpretação e avaliação dos resultados.
- **Resultados sem Combinação:** neste capítulo são apresentados os experimentos realizados com as representações individualmente nesse trabalho.
- **Experimentos Computacionais:** neste capítulo são apresentados os experimentos realizados com as representações combinadas. Para cada experimento realizado é apresentado: a combinação dos dados utilizados, as técnicas testadas, os resultados obtidos, e a interpretação e análise dos resultados.
- **Considerações Finais:** neste capítulo são apresentadas as considerações finais do trabalho realizado e os trabalhos futuros.

2 PROBLEMA

A classificação automática de gêneros musicais é uma tarefa importante considerando o crescimento contínuo dos dados musicais nos últimos anos (algumas Bibliotecas de Músicas Digitais possuem milhões de músicas) e também pela popularidade dos gêneros para a categorização das músicas (FABBRI, 1999a; LEE; DOWNIE, 2004; ORIO, 2006; MCKAY; FUJINAGA, 2006; NEUMAYER; RAUBER, 2007). Para a tarefa de classificação automática de gêneros musicais é possível usar dados culturais (WHITMAN; SMARAGDIS, 2002), áudio (LI et al., 2003; SCARINGELLA et al., 2006; FU et al., 2011; SILLA JR. et al., 2010) e/ou letras de músicas (FELL; SPORLEDER, 2014). Esse trabalho foca na classificação automática de gêneros musicais utilizando letras de músicas, mas com potencial para auxiliar e em alguns casos melhorar os resultados da classificação de gêneros musicais utilizando áudio, tendo em vista que o uso de áudio isoladamente atingiu um “limite” (“*glass ceiling*”) (PACHET; AUCOUTURIER, 2004), podendo inclusive ser utilizada individualmente (MAYER et al., 2008a, 2008b; MAYER; NEUMAYER, 2009; MAYER; RAUBER, 2010, 2011). As letras de músicas possuem estrutura e vocabulário diferentes de textos comuns, desta forma as técnicas padrões utilizadas na classificação de texto não necessariamente obtém os mesmos resultados na classificação de gêneros musicais utilizando letras de músicas. Além disso, as ferramentas PLN são desenvolvidas majoritariamente para o idioma Inglês (MAYER et al., 2008a, 2008b; MCKAY et al., 2010; MAYER; RAUBER, 2011; FELL; SPORLEDER, 2014).

Como as técnicas de PLN são desenvolvidas tendo como foco principal o idioma Inglês, muitas vezes a aplicação de tais técnicas em outros idiomas se torna inviável, pois cada idioma tem estrutura e vocabulário específicos. Note que a motivação para aplicação de técnicas em diversos idiomas é relevante, pois a música é um fenômeno global e está presente em diversos países e culturas.

Além dos trabalhos da literatura focarem principalmente no desenvolvimento de técnicas para o idioma Inglês (MAYER et al., 2008a, 2008b; MCKAY et al., 2010; MAYER;

RAUBER, 2011; FELL; SPORLEDER, 2014), muitas vezes estas técnicas são exclusivamente dependentes de idioma. Desta forma, as técnicas desenvolvidas acabam considerando características exclusivas do idioma em questão, como por exemplo, gírias e estruturas ortográficas. Neste contexto, torna-se difícil inclusive adaptar o uso para outros idiomas.

3 REFERENCIAL TEÓRICO

Neste capítulo são abordadas as ferramentas, técnicas e recursos utilizados para a mineração de texto conforme a Seção 3.1. Na Seção 3.2 são apresentados os algoritmos e métricas de avaliação utilizados no aprendizado supervisionado de máquina. Na Seção 3.3 são apresentadas as técnicas de combinação de classificadores: *early* e *late fusion*. Na Seção 3.4 são apresentados os trabalhos relacionados com classificação automática de gêneros musicais utilizando letras de músicas.

3.1 MINERAÇÃO DE TEXTOS

A Mineração de Textos é uma área da inteligência artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado e também a construção de sistemas inteligentes capazes de obter conhecimento de forma automática em grandes quantidades de dados (MONARD; BARANAUSKAS, 2003; LAROSE, 2014). Nesta seção são apresentadas algumas das técnicas comumente utilizadas em Mineração de Textos para extrair características e padronizar os dados: Na Seção 3.1.1 é apresentado o processo de remoção de *stopwords*; A técnica *stemming* é apresentada na Seção 3.1.2; Uma abordagem alternativa são os n-gramas apresentados na Seção 3.1.3; O TF-IDF é utilizado em conjunto com as abordagens anteriores para se obter a importância das palavras, o qual é apresentado na Seção 3.1.4; por fim, na Seção 3.1.5 são apresentadas as características estilísticas.

3.1.1 REMOÇÃO DE *STOPWORDS*

As *stopwords* são palavras comuns (por exemplo, artigos e preposições) que carregam pouco significado semântico para o texto. Estas palavras irrelevantes são incluídas em uma lista pré-definida chamada “lista de *stopwords*” ou “*stoplist*”.

3.1.2 STEMMING

A técnica *stemming* é utilizada para obter o radical de todas as palavras a partir de um conjunto de textos. Por exemplo, dadas as palavras “*computing*” e “*computer*” em Inglês, o resultado do algoritmo de *stemming* poderia ser “*comput*” (PORTER, 1980). Sebastiani (SEBASTIANI, 2002) mostra que esta técnica é largamente utilizada no domínio de classificação de textos com o objetivo de reduzir a dimensionalidade. Este também é um procedimento muito rápido. É importante ressaltar que os algoritmos de “*stemming*” são dependentes de linguagem e, portanto podem não estar disponíveis para alguns idiomas (WU; SALTON, 1981).

3.1.3 CARACTERES N-GRAMAS

Esta técnica transforma palavras em *n*-gramas. Um *n*-grama é uma sequência de *n* caracteres (adicionando o caractere especial “_” para denotar o início e o final da palavra), por exemplo, dado o termo “*text*”: para $n = 2$, os atributos representativos poderiam ser “_t”, “te”, “ex”, “xt” e “t_”; para $n = 3$, “_te”, “tex”, “ext” e “xt_”; e para $n = 4$, “_tex”, “text” e “ext_” (CAVNAR; TRENKLE, 1994). Note que a abordagem por caracteres *n*-gramas é ligeiramente diferente das *word n*-gramas, as quais são palavras utilizadas em conjunto para formar os *n*-gramas. Por exemplo, na abordagem por *word n*-gramas com $n = 2$, o texto “mineração de texto em classificação de gêneros musicais” gera os seguintes atributos: “mineração de”, “de texto”, “texto em”, “em classificação”, “classificação de”, “de gêneros” e “gêneros musicais”. É importante ressaltar que as *word n*-gramas são utilizadas em idiomas orientais devido a dificuldade de se separar palavras em letras em tais idiomas (HU et al., 2009).

Em casos distintos de idiomas esta técnica é geralmente utilizada quando os idiomas tem um radical comum. Desta forma, a variação de algumas palavras não influenciam os radicais que são comuns entre os idiomas. Portanto, esta abordagem permite inferir morfemas dividindo a palavra em uma sequência de tamanho fixo, *n*, a qual é pré definida de acordo com o contexto da aplicação. Note que os *n*-gramas gerados são únicos, portanto, não há repetição nas características geradas.

3.1.4 TF-IDF

Existem muitos métodos para calcular o peso de cada termo (por exemplo, *Weighted Inverse Document Frequency* (TOKUNAGA; MAKOTO, 1994) e Relevância do Termo (WU; SALTON, 1981)), sendo o TF-IDF um dos mais comumente empregados (SALTON; BUC-

KLEY, 1988). A vantagem do TF-IDF é o *trade-off* entre quanto um termo ocorre (frequência do termo) e em quantos documentos o termo ocorre. Significando que um termo que ocorre frequentemente em um dado documento pode ser importante, mas somente se este não ocorrer em todos os outros documentos. A fórmula para computar o TF-IDF é apresentada na Equação (1).

$$TFIDF = TF \times \log \frac{nD}{nDF} \quad (1)$$

onde:

- TF é a frequência de um termo no documento, em outras palavras, o número de vezes que o termo ocorre em um documento;
- nD é o número de documentos utilizados;
- nDF é o número de documentos onde ocorre a representação do termo;

3.1.5 CARACTERÍSTICAS ESTILÍSTICAS DE TEXTO

A utilização das características estilísticas de texto para a classificação de gêneros musicais baseada em letras de músicas foi originalmente proposta no trabalho de (MAYER et al., 2008b). No trabalho de (MAYER et al., 2008b) eles propuseram contagens simples de pontuações especiais (por exemplo, “!”, “-”, “?”), ocorrência de dígitos, palavras e palavras únicas por linha, taxa de palavras únicas, caracteres por palavra e palavras por linha. A principal motivação de utilizar as características estilísticas é captar as “palavras de interjeição” nas letras de músicas, pontuações especiais e computar estatísticas simples de texto. Posteriormente, estas características estilísticas de texto foram complementadas com novas métricas propostas (conforme Tabela 1) para a tarefa de classificação de emoções baseada em letras de músicas (HU; DOWNIE, 2010a). A utilização das características estilísticas de texto para a classificação de gêneros musicais baseada em letras de músicas foi originalmente proposta no trabalho de (MAYER et al., 2008b).

3.2 APRENDIZADO SUPERVISIONADO

De acordo com (MONARD; BARANAUSKAS, 2003) um algoritmo baseado no aprendizado supervisionado, recebe um conjunto de exemplos de treinamento. Geralmente esses

Tabela 1: Características estilísticas

Característica	Definição
palavras de interjeição	freq. normal. “ja”, “hejsan”, “yo”, “uh”, “ah”, “aj”, “yeah”, “hej”, “vad”, “oj”, “hey”, “hei”, “ooh”
pontuações especiais	frequência normalizada de “!”, “:”, “?”, “-”, “;”
NÚMERO	frequência normalizada dos números
Número de palavras	número total de palavras
Número de palavras únicas	número total de palavras únicas
taxaDePalavrasRepetidas	$(N^{\circ} \text{ de palav.} - N^{\circ} \text{ de palav. únicas}) / N^{\circ} \text{ de palav.}$
Média de tamanho da palavra	número médio de caracteres por palavra
Número de linhas	número total de linhas
Número de linhas únicas	número total de linhas únicas
Número de linhas em branco	número de linhas em branco
taxaDeLinhasEmBranco	$N^{\circ} \text{ de linhas em branco} / N^{\circ} \text{ de linhas}$
médiaDeTamanhoDeLinhas	$N^{\circ} \text{ de palavras} / N^{\circ} \text{ de linhas}$
desvioPadraoDoTamanhoDaLinha	desvio padrão do número de palavras por linha
palavrasÚnicasPorLinha	$N^{\circ} \text{ de palavras únicas} / N^{\circ} \text{ de linhas}$
taxaDeLinhaRepetida	$(N^{\circ} \text{ de linhas} - N^{\circ} \text{ de linhas únicas}) / N^{\circ} \text{ de linhas}$
taxaMédiaDePalavraRepetidaPorLinha	taxa média de palavra repetida por linha
taxaDesvPadraoPalavRepetPorLinha	taxa de desvio padrão de palavra repetida por linha

Fonte: Adaptado de (HU; DOWNIE, 2010a)

exemplos são compostos por um vetor de características e o rótulo da classe associada. Nesse caso, o objetivo do algoritmo é construir um classificador, com base nas amostras de exemplo, que seja capaz de rotular uma nova amostra sem rótulo que seja inserida posteriormente. Em síntese, esse tipo de aprendizado é utilizado quando as classes já são conhecidas pelo usuário. Na seção 3.2.1 é apresentado o classificador supervisionado *Support Vector Machines* (SVM) e na seção 3.2.2 são apresentados os procedimentos utilizados para avaliar os resultados obtidos pelo SVM.

3.2.1 CLASSIFICADOR SVM

O SVM é um algoritmo de aprendizado supervisionado de máquina que utiliza hiperplanos para separar duas classes através do mapeamento dos exemplos de treinamento (BURGES, 1998). Com o objetivo de classificar múltiplas classes foi utilizado o algoritmo¹ *Sequential Minimal Optimization* (SMO) para o treinamento com os parâmetros padrões e as características normalizadas (PLATT, 1999).

¹Implementação da ferramenta WEKA (HALL et al., 2009)

O treinamento de um SVM requer a solução de um problema de otimização de programação quadrática (PLATT, 1999). O SMO divide este problema de programação quadrática em uma série de problemas menores. Estes problemas de programação quadrática menores são resolvidos analiticamente usando otimização numérica em repetições internas. A quantidade de memória necessária para o SMO é linear no conjunto de treinamento, a qual permite ao SMO manipular grandes conjuntos. Desta forma, o SMO é mais rápido do que o SVM para conjuntos de dados lineares e esparsos.

Em (DHANARAJ; LOGAN, 2005) foi demonstrado que os SVMs apresentam três principais vantagens: não requer um ajuste complexo dos parâmetros, pode generalizar pequenos conjuntos e são adequados para o aprendizado em espaços com alta dimensionalidade. Além disso, o classificador SVM tem recebido atenção em MIR principalmente pelo desempenho e melhores resultados comparado a outros classificadores padrões como Árvore de Decisão, K-Vizinhos Mais Próximos e *Random Forests* (LAURIER et al., 2008). Entre outros artigos que usaram o SVM e outras derivações para a tarefa de classificação de gêneros musicais estão (MAYER et al., 2008a, 2008b; MAYER; NEUMAYER, 2009; MAYER; RAUBER, 2010, 2011), portanto, optou-se por utilizar o SVM como classificador devido a maioria dos trabalhos de classificação de gêneros musicais utilizando letras de músicas obterem os melhores resultados através do SVM, devido a sua capacidade de lidar com espaços de dimensionalidade em grande escala.

3.2.2 MÉTODOS PARA A AVALIAÇÃO DE MODELOS

3.2.2.1 VALIDAÇÃO CRUZADA

De acordo com (MONARD; BARANAUSKAS, 2003; LAROSE, 2014) este é um tipo de abordagem baseada em amostragem. A validação cruzada divide aleatoriamente o conjunto de dados em r partições mutuamente exclusivas. Essas partições possuem tamanho aproximadamente igual a n/r amostras. Os exemplos nas partições ($r - 1$) são utilizados para treinamento e a hipótese é testada na partição remanescente. Este processo é repetido r vezes, cada uma considerando uma partição diferente para teste. O erro é estimado por meio da média dos erros calculados em cada uma das r partições.

3.2.2.2 MÉTRICAS PARA AVALIAÇÃO DE RESULTADOS

Para avaliar o desempenho da classificação podem ser utilizadas as seguintes métricas: precisão, que calcula a porcentagem de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas (Equação 2); sensibilidade, a qual retorna a porcentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas (cobertura conforme Equação 3); e *f-measure*, também chamada de *f-score*, é a média ponderada de precisão e sensibilidade (Equação 4) (KOHAVI; PROVOST, 1998). Vp indica o número de exemplos positivos da classe que foram previstos corretamente, Fp refere-se ao número de exemplos positivos da classe que foram classificados incorretamente e Fn que indica o número de exemplos negativos da classe que foram previstos incorretamente.

$$Precisão = \frac{Vp}{Vp + Fp} \quad (2)$$

$$Cobertura = \frac{Vp}{Vp + Fn} \quad (3)$$

$$F\text{-measure} = 2 \cdot \frac{Precisão \times Cobertura}{Precisão + Cobertura} \quad (4)$$

3.3 COMBINAÇÃO DE CLASSIFICADORES

Neste trabalho foram feitas diversas combinações das representações individuais para verificar se a fusão dos métodos melhora os resultados da classificação. Estas combinações foram feitas utilizando os métodos *early fusion* (Seção 3.3.1) e *late fusion* (Seção 3.3.2).

3.3.1 EARLY FUSION

A Figura 1 ilustra a técnica *early fusion*. De acordo com (SNOEK et al., 2005) *early fusion* é o esquema de fusão que integra características unimodais antes da etapa de aprendizado. Desta forma, abordagens de indexação que são baseadas em *early fusion* primeiramente extraem as características unimodais. Depois da análise de várias características de texto, as características extraídas são combinadas em uma representação simples. Após a combinação das características unimodais de texto em uma representação multimodal, os métodos *early fusion* iniciam o aprendizado supervisionado para classificar baseado nos conceitos semânticos.

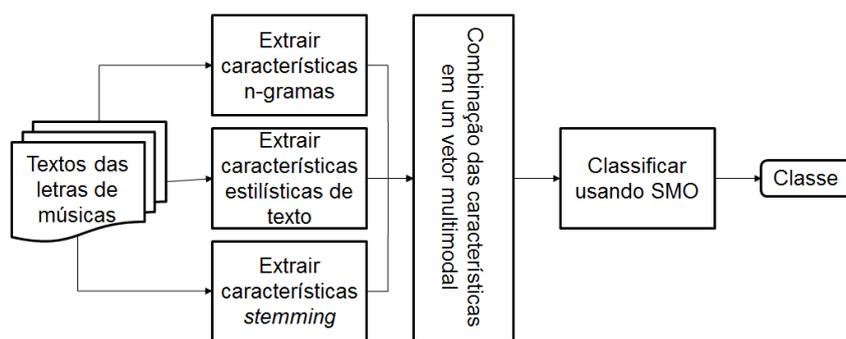


Figura 1: Esquema de combinação de características por *early fusion*.

Fonte: Autoria própria

Uma vantagem adicional é a necessidade de uma fase de aprendizado apenas. A desvantagem desta abordagem é a dificuldade para combinar características em uma representação comum. A técnica de *early fusion* foi previamente aplicada em (SNOEK et al., 2005; LIMA et al., 2014).

3.3.2 LATE FUSION

De acordo com (SNOEK et al., 2005) *late fusion* é um esquema de fusão que reduz as características unimodais para realizarem o aprendizado separadamente, posteriormente integrando as saídas. Desta forma, abordagens de indexação que são baseadas em *late fusion* também iniciam com a extração das características unimodais. Em contraste com *early fusion*, onde as características são combinadas em representação multimodal, abordagens baseadas em *late fusion* realizam o aprendizado a partir das características unimodais. Em geral, esquemas *late fusion* combinam as saídas dos classificadores unimodais em uma representação multimodal.

Late Fusion foca nas características individuais de cada modalidade. As saídas unimodais são fundidas em uma representação semântica multimodal. Uma grande desvantagem de esquemas *late fusion* é a performance que acaba sendo comprometida, pois muitas modalidades requerem um estágio de aprendizado supervisionado separadamente. Outra desvantagem das abordagens *late fusion* é a potencial perda de correlação no espaço de características mixada.

O uso da combinação de classificadores (uma forma de uso dos métodos *late fusion*) tem por objetivo melhorar a performance de classificação através da complementariedade das saídas dos classificadores. Neste trabalho foram usados 3 métodos de combinação de classifica-

dores, sendo estes os mais comuns encontrados na literatura: probabilidade máxima, soma das probabilidades e produto da probabilidade (KITTLER et al., 1998). Neste sentido, foi usada uma abordagem similar na qual é possível usar um classificador com diferentes representações combinando as saídas (para cada representação é gerado um modelo de treinamento). Esta abordagem é conhecida também como *late fusion* (XU et al., 1992; HO et al., 1994).

3.3.2.1 PROBABILIDADE MÁXIMA

A partir das probabilidades obtidas de cada classificador é realizado o procedimento para encontrar uma classe com a maior probabilidade entre as representações sendo esta a saída gerada da combinação de classificadores. A definição formal apresentada na Equação (5) é baseada em (KITTLER et al., 1998):

$$\begin{aligned} & \text{classifique } Z \rightarrow w_j \text{ se} \\ & P(w_j|x_1, \dots, x_R) = \max_k P(w_k|x_k, \dots, x_R), \end{aligned} \quad (5)$$

onde Z é uma instância para ser classificada em uma das m classes possíveis (w_1, \dots, w_m). Assuma que R classificadores (neste caso R representa modelos de representações) denota um vetor usado pelo i th classificador x_i . A Figura 2 ilustra a técnica *late fusion* usando probabilidade máxima.

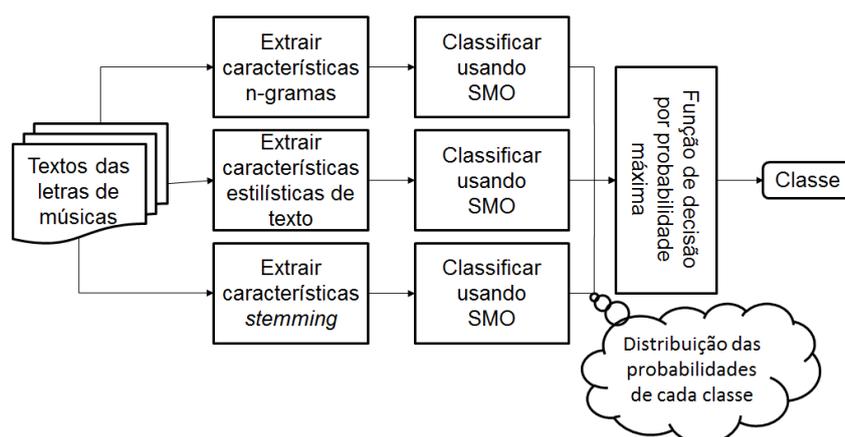


Figura 2: Esquema de combinação de características por *late fusion* usando probabilidade máxima.

Fonte: Autoria própria

3.3.2.2 SOMA DAS PROBABILIDADES

Este método faz a soma da probabilidade de cada rótulo obtido pelos classificadores e seleciona a classe com o valor mais alto. Desta forma, cada classificador auxilia individualmente a saída da classificação e colabora para uma decisão consensual tornando a classificação mais consistente com a melhoria da acurácia. Seguindo as definições em (KITTLER et al., 1998) a equação (6) demonstra as definições formais para obter a saída do método da soma da probabilidade. A Figura 3 ilustra a técnica *late fusion* usando soma das probabilidades.

$$\text{classifique } Z \rightarrow w_j \text{ se} \quad (6)$$

$$P(w_j) + \sum_{i=1}^R P(w_j|x_i) = \max_{k=1}^m \left[P(w_k) + \sum_{i=1}^R P(w_k|x_i) \right]$$

onde Z é uma instância para ser classificada em uma das m classes possíveis (w_1, \dots, w_m). Assuma que R classificadores (neste caso R representa modelos de representações) denota um vetor usado pelo i th classificador x_i .

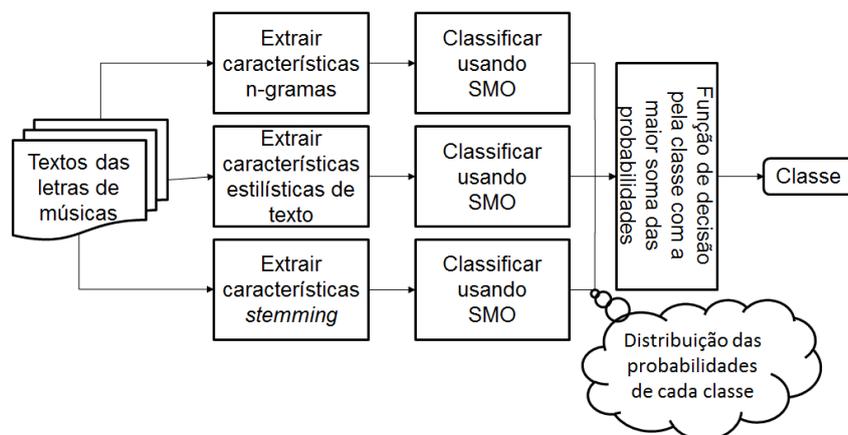


Figura 3: Esquema de combinação de características por *late fusion* usando soma das probabilidades.

Fonte: Autoria própria

3.3.2.3 PRODUTO DAS PROBABILIDADES

A decisão baseada no produto da probabilidade é realizada pela multiplicação entre as probabilidades das classes retornada de cada representação ou classificador. Então, é selecionada a classe com o valor mais alto. Esta abordagem considera que as representações usadas

são condicionalmente independentes. Seguindo as definições em (KITTLER et al., 1998) a Equação (7) demonstra as definições formais para obter a saída do método do produto da probabilidade. A Figura 4 ilustra a técnica *late fusion* usando o produto das probabilidades.

$$\text{classifique } Z \rightarrow w_j \text{ se}$$

$$P(w_j) + \prod_{i=1}^R P(w_j|x_i) = \max_{k=1}^m P(w_k) + \prod_{i=1}^R P(w_k|x_i) \quad (7)$$

onde Z é uma instância para ser classificada em uma das m classes possíveis (w_1, \dots, w_m). Assuma que R classificadores (neste caso R representa modelos de representações) denota um vetor usado pelo i th classificador x_i .

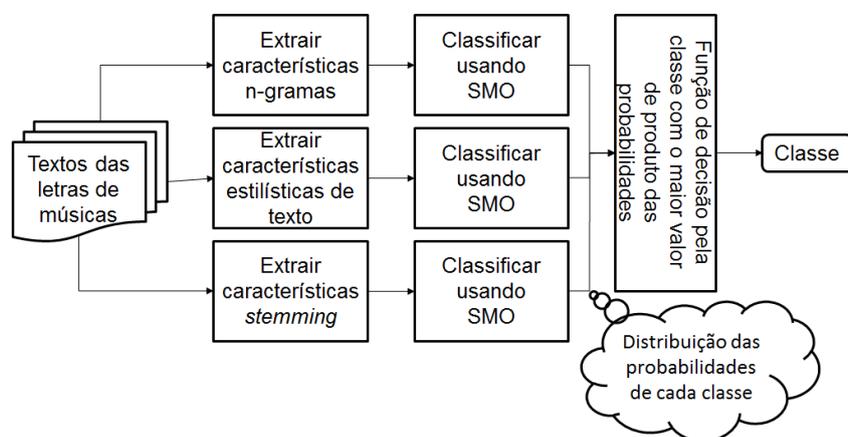


Figura 4: Esquema de combinação de características por *late fusion* usando produto das probabilidades.

Fonte: Autoria própria

3.4 TRABALHOS RELACIONADOS

No trabalho de (MAYER et al., 2008b) os autores propuseram o uso de características de rimas, estilísticas e *Part-Of-Speech* (POS) que é uma técnica de categorização de palavras com propriedades gramaticais similares. Além disso, em (MAYER et al., 2008b) foi proposta a combinação dessas características com a abordagem de *Bag Of Words* (BOW) ponderadas com *Term Frequency Inverse Document Frequency* (TF-IDF) para a tarefa de classificação de gêneros musicais baseada em letras de músicas. Nos experimentos de (MAYER et al., 2008b) foi utilizada uma base de dados com 397 letras de músicas em Inglês distribuídas em 10 gêneros musicais (*Country*: 29, *Folk*: 44, *Grunge*: 40, *Hip Hop*: 41, *Metal*: 46, *Pop*: 42, *Punk Rock*:

40, *R&B*: 40, *Reggae*: 33, *Slow Rock*: 42). O melhor resultado obtido nos experimentos de (MAYER et al., 2008b) foi utilizando o classificador SVM com um vetor combinado das características de rimas e estilísticas com POS. Eles também reportaram bons resultados utilizando as características estilísticas isoladamente. Porém, neste trabalho foram utilizadas técnicas específicas de idioma, como por exemplo, as características de rimas, as quais são dependentes de idioma. Além disso, os autores utilizaram uma base de dados exclusivamente com letras de músicas em Inglês, sendo que não necessariamente poderão ser obtidos os mesmos resultados em bases de dados com letras em outros idiomas.

No trabalho de (MAYER et al., 2008a) os autores propuseram combinar as características de rimas e estilísticas desenvolvidas anteriormente (baseadas em letras) com características padrões baseadas em áudio para a tarefa de classificação de gêneros musicais. Nos experimentos de (MAYER et al., 2008a) foram empregadas duas bases de dados. A primeira base de dados (*collection_600*) tem 600 músicas uniformemente distribuídas entre 10 gêneros musicais. A segunda base de dados (*collection_3010*) tem 3.010 músicas dos mesmos gêneros musicais, mas com a seguinte distribuição: *Country*: 227, *Folk*: 179, *Grunge*: 181, *Hip Hop*: 381, *Metal*: 371, *Pop*: 371, *Punk Rock*: 374, *R&B*: 373, *Reggae*: 181, *Slow Rock*: 372. Nos experimentos de (MAYER et al., 2008a), o melhor resultado para a primeira base de dados (*collection_600*) foi obtido com o classificador SVM usando características de áudio conhecidas como *Statistical Spectrum Descriptors* (SSD) em combinação com as características estilísticas de texto e POS (extraídas de letras) através da técnica de *early fusion*. Para a segunda base de dados (*collection_3010*) o melhor resultado foi também obtido pelo uso de SSD em combinação com características estilísticas de texto e POS através da técnica de *early fusion*.

A mesma abordagem usada em (MAYER et al., 2008a) foi estendida (MAYER; NEUMAYER, 2009) para lidar com uma base de dados de 20.109 músicas com 14 gêneros musicais (*Pop*: 6.156, *Alternative*: 3.699, *Rock*: 3.666, *Hip Hop*: 2.234, *Country*: 1.990, *R&B*: 1.107, *Christian*: 490, *Comedy*: 206, *Reggae*: 121, *Dance/Electronic*: 112, *Blues*: 99, *Jazz*: 97, *Scores/Soundtrack*: 70, *Classical*: 62). O melhor resultado obtido por (MAYER et al., 2008a) foi através da combinação de SSD com todas as características baseadas em letras de músicas.

No trabalho de (MCKAY et al., 2010) foi avaliada a performance de classificação de gêneros musicais utilizando características baseadas em letras de músicas em relação a características de áudio, simbólicas ou culturais. Nos experimentos foi utilizada uma base de dados de 250 músicas com letras em Inglês distribuídas igualmente entre 10 gêneros musicais (*Modern Blues*, *Traditional Blues*, *Baroque*, *Romantic*, *Bop*, *Swing*, *Hardcore Rap*, *Pop*

Rap, *Alternative Rock*, *Metal*). Na avaliação realizada por (MCKAY et al., 2010) os resultados obtidos demonstraram que as características baseadas em letras de músicas frequentemente melhoram os resultados de classificação ao serem combinadas com outras abordagens através da técnica de *early fusion*.

No trabalho de (FELL; SPORLEDER, 2014), novas características baseadas em letras de músicas foram propostas. Mais precisamente, os autores definiram 13 conjuntos de características baseadas em letras de músicas que se enquadram em uma de cinco dimensões: Vocabulário, Estilo, Semântica, Orientação ou Estrutura da Música. A dimensão do vocabulário é composta dos *top K* n-gramas com TF-IDF e palavras sem padrão (por exemplo, palavras incomuns ou gírias). A dimensão estilo é composta de POS, *echoism* (repetições de letras ou palavras, por exemplo, “*money, money*”) e características de rimas. Note que as características de rimas de (FELL; SPORLEDER, 2014) são baseadas exclusivamente no gênero musical *Rap* do trabalho de (HIRJEE; BROWN, 2010), diferindo portanto das demais utilizadas em outros trabalhos (MAYER et al., 2008b, 2008a; MAYER; NEUMAYER, 2009) que são baseadas em outros gêneros. As características na dimensão semântica foram obtidas com base nos trabalhos de (MARTINDALE, 1975, 1990). Para a dimensão de orientação eles usam a informação se a canção narra experiências sobre o passado ou presente/futuro. Eles também modelaram como “egocêntrica” uma música pela computação de frequências de pronomes. Para a dimensão de estrutura da música eles pesquisaram por estruturas repetitivas nas letras de músicas. Nos experimentos de (FELL; SPORLEDER, 2014) foi empregada uma base de dados com 3.360 músicas com letras em Inglês de 8 gêneros musicais diferentes (*Blues*, *Rap*, *Metal*, *Folk*, *R&B*, *Reggae*, *Country* e *Religious*). O melhor resultado obtido com esses experimentos foi utilizando o classificador SVM com todos os 13 conjuntos de características através de *early fusion* baseadas em letras de músicas.

A análise dos trabalhos relacionados apresentada nesta subseção mostra que para o problema de classificação de gêneros musicais baseada em letras de músicas, somente letras no idioma Inglês tem sido usadas. Infelizmente a maioria das características propostas baseadas em letras de músicas não podem ser usadas devido a falta de recursos linguísticos compatíveis. Uma das exceções é o uso das características estilísticas introduzidas em (MAYER et al., 2008b) e que tem sido utilizadas neste trabalho. Além disso, a ideia de combinação da representação BOW (tanto utilizando a abordagem *stemming* ou caracter n-gramas) em combinação somente com as características estilísticas têm sido negligenciada. Na Tabela 2 são descritos resumidamente os principais trabalhos do estado da arte.

Tabela 2: Resumo dos trabalhos relacionados

Trabalho	Características	Base(s)	Melhor Resultado
Mayer et al. (2008b)	Rimas, Estilísticas, POS, BOW, TFIDF	397 letras em Inglês; 10 gêneros	SVM+ junção das características
Mayer et al. (2008a)	SSD+Mesmas de (MAYER et al., 2008b)	Collection_600/ collection_3010; 10 gêneros	SVM+ todas as características
Mayer e Neumayer (2009)	Mesmas de (MAYER et al., 2008a)	20.109 músicas; 14 gêneros	Junção das características
McKay et al. (2010)	Letras, simbólicas, áudio, culturais	250 músicas em Inglês; 10 gêneros	Junção das características
Fell e Sporleder (2014)	13 conjuntos em 5 dimensões: estilo, vocabulário, semântica, orientação ou estrutura	3.360 letras em Inglês; 8 gêneros	SVM+ junção das características

4 METODOLOGIA

Neste capítulo são apresentadas as etapas realizadas para a execução dos experimentos: Na Seção 4.1 são apresentados os detalhes das bases de dados criadas; e na Seção 4.2 são apresentadas as técnicas utilizadas para o processamento das letras de músicas.

4.1 CRIAÇÃO DAS BASES DE DADOS

Com o intuito de atender ao objetivo principal deste trabalho, ou seja, analisar a eficácia de diferentes abordagens para a classificação automática de gêneros musicais usando letras de músicas em diferentes idiomas, foi necessário criar novas bases de dados para esse propósito, pois como apresentado no capítulo 2, os trabalhos existentes focam exclusivamente na classificação de gêneros musicais em inglês.

4.1.1 NORDIC LYRICS DATABASE

A *Nordic Lyrics Database* (NLD) foi criada em (LIMA et al., 2014) usando o seguinte procedimento: no primeiro estágio foi aplicado um questionário para alguns residentes da Dinamarca, Noruega e Suécia. Neste questionário os participantes foram instruídos para informar as músicas favoritas com os respectivos títulos, o artista, gênero musical e idioma. Note que foi imposta uma restrição no questionário para informar apenas músicas da Dinamarca, Noruega e Suécia. No segundo estágio, as músicas foram buscadas manualmente na Web para recuperar as letras para os experimentos deste trabalho.

A NLD contém 1.513 letras de músicas divididas em cinco gêneros (Dance: 265, Pop Rock: 263, Rap: 100, Pop: 357 and Rock: 528) distribuídos em três idiomas Nórdicos (Sueco, Dinamarquês e Norueguês). Os gêneros musicais foram rotulados seguindo a definição demonstrada de (FABBRI, 1999b): “*a kind of music, as it is acknowledged by a community for any reason or purpose or criteria*”. A NLD contém alguns termos no idioma Inglês conforme

Português e Inglês). Nas Figuras 10-19 são apresentadas as nuvens de palavras com os 150 maiores valores de TFIDF para cada gênero da LMLD.



Figura 10: Nuvem de palavras para o gênero axé da LMLD.

Fonte: Autoria própria

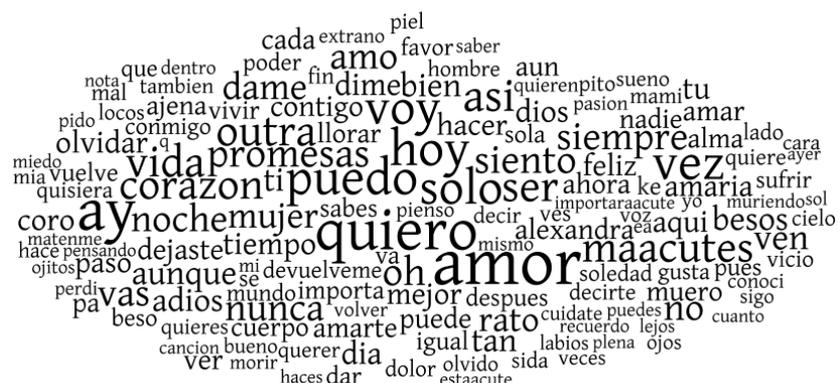


Figura 11: Nuvem de palavras para o gênero bachata da LMLD.

Fonte: Autoria própria



Figura 12: Nuvem de palavras para o gênero bolero da LMLD.

Fonte: Autoria própria

diferença nos atributos para algumas representações (por exemplo, 2-gramas para a LMLD), existe diferença nos arquivos de características gerados, os quais podem gerar diferentes resultados no experimentos reportados. Note também que não foram aplicados métodos de redução de dimensionalidade devido aos trabalhos da literatura não realizarem tais tarefas e também a experimentos anteriores retornarem resultados inferiores utilizando métodos como: características com melhor TF-IDF ou com melhor TF.

Tabela 3: Número de atributos para cada representação sem e com remoção de *stopwords* para cada base de dados de letras de músicas.

Abordagem	LGTZAN	LMLD	NLD	JLD
<i>Stemming</i> -SWR	7.913	22.087	20.962	48.658
<i>Stemming</i> +SWR	7.850	22.025	20.704	46.201
2-Gramas-SWR	681	860	1.087	1.103
2-Gramas+SWR	680	860	1.079	1.103
3-Gramas-SWR	4.790	6.330	9.945	10.942
3-Gramas+SWR	4.781	6.321	9.829	10.942
4-Gramas-SWR	13.707	24.778	36.741	49.359
4-Gramas+SWR	13.653	24.741	35.705	49.326

Fonte: Autoria própria

4.2 PROCESSAMENTO DE LETRAS DE MÚSICAS

Uma abordagem comumente empregada para o processamento de texto é o BOW (SEBASTIANI, 2002), no qual os documentos são representados em vetores de palavras que ocorrem nos documentos. Esta abordagem é necessária para padronizar as letras de músicas em um formato fixo e compreendido pelo classificador. Desta forma, nesta seção são descritos os métodos utilizados no processamento das letras de músicas e também na fase de pré-processamento, como mostrado na Figura 28. Na fase de pré-processamento são realizados os seguintes passos: identificação do idioma, remoção dos caracteres especiais e a padronização das letras de músicas. Existem 5 etapas independentes na fase de processamento das letras de músicas, são elas: processo opcional de remoção das *stopwords* (note que embora as *stopwords* sejam dependentes de idiomas foram utilizadas em ambas as abordagens para complementar os experimentos), o *stemming* (dependente de idioma), os caracteres n-gramas (que é independente de idioma), o método TF-IDF (utilizado com *stemming* e n-gramas) e as características estilísticas de texto.

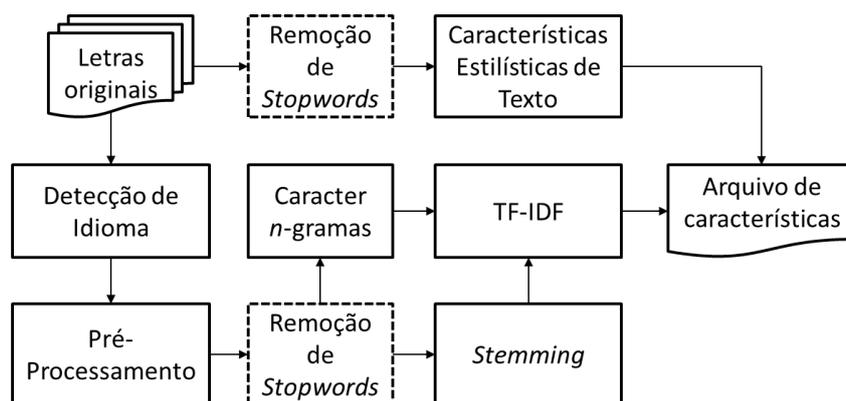


Figura 28: Visão geral das etapas de processamento de letras de músicas.

Fonte: Autoria própria

Tabela 4: Mapeamento de caracteres especiais para os idiomas Nórdicos

Idioma(s)	Caracteres Especiais	Caracter equivalente
Dinamarquês, Norueguês	æ	ae
Dinamarquês, Norueguês	ø	oe
Dinamarquês, Norueguês, Sueco	å	aa
Sueco	ö	oe
Sueco	ä	ae

Fonte: Autoria própria.

4.2.1 DETECÇÃO DE IDIOMAS

Esta etapa de processamento de letras de músicas consiste na detecção do idioma do documento utilizando uma Biblioteca de código aberto em linguagem Java chamada “*Language Detection*” (SHUYO, 2010).

4.2.2 PRÉ-PROCESSAMENTO

Antes do processamento das letras de músicas é importante ter uma fase de pré-processamento para remoção de caracteres especiais, pois existem algumas letras de músicas que precisam ser padronizadas para manter seu significado (idiomas Nórdicos), conforme a Tabela 4. A próxima etapa consiste em colocar todas as palavras em minúsculo para auxiliar as comparações entre as palavras. Depois desta etapa de pré-processamento serão obtidos documentos padronizados sem símbolos inúteis.

4.2.3 *STOPLISTS* UTILIZADAS

Neste trabalho são utilizadas uma “*stoplist*”¹ para cada idioma, sendo: 114 para Sueco, 176 para Norueguês e 94 para Dinamarquês.

4.2.4 CARACTERÍSTICAS ESTILÍSTICAS DE TEXTO UTILIZADAS

Neste trabalho foi empregado um subconjunto de características estilísticas propostas para a tarefa de classificação de emoções baseada em letras de músicas (HU; DOWNIE, 2010a). Este subconjunto foi utilizado porque complementa as características estilísticas originais com novas métricas conforme apresentado anteriormente na seção 3.1.5 e também devido a melhora de resultados obtida em (MAYER et al., 2008b). Neste trabalho não foram usadas as características estilísticas relacionadas aos arquivos de áudio (por exemplo, número de palavras por minuto ou o número de linhas por minuto). A principal vantagem das características estilísticas é que são computadas medidas estatísticas simples baseadas na frequência da palavra ou caractere.

¹As “*stoplists*” foram obtidas do endereço <http://snowball.tartarus.org>

5 RESULTADOS SEM COMBINAÇÃO (*NO FUSION*) PARA OS EXPERIMENTOS COMPUTACIONAIS

Neste capítulo serão respondidas as seguintes questões através de experimentos controlados: Quando lidamos com cada conjunto de dados que tem letras de músicas em diferentes idiomas qual representação deveria ser utilizada? Uma representação dependente de idioma (usando ferramentas de idioma tais como *stemmers* ou remoção de *stopwords*) é melhor do que uma representação independente de idioma (usando n-gramas)? Qual é o impacto das *stopwords* na classificação de gêneros musicais utilizando letras de músicas? Além disso, o que acontece com todas essas questões quando os três conjuntos de dados são combinados na JLD? Para avaliar os resultados neste trabalho foi usada a *F1-score* (a qual é a média harmônica das métricas de *precision* e *recall*) com o procedimento de validação cruzada fator 10.

5.1 RESULTADOS PARA ABORDAGENS SEM COMBINAÇÃO (*NO FUSION*)

Os resultados (em %) detalhados sem (-SWR) ou com (+SWR) remoção de *stopwords* para cada representação *no fusion* são apresentados na Tabela 5.

Tabela 5: Resultados para cada base de dados individualmente

Abordagem	LGTZAN	LMLD	NLD	JLD
<i>Stemming</i> -SWR	34.3	37.6	38.7	35.3
<i>Stemming</i> +SWR	40.4	31.7	35.1	30.9
2-Gramas-SWR	34.6	51.3	62.9	46.7
2-Gramas+SWR	36.7	49.9	62.6	46.1
3-Gramas-SWR	42.6	60.3	77.4	51.8
3-Gramas+SWR	40.9	60.3	78.4	52.2
4-Gramas-SWR	43.9	65.6	87.1	49.7
4-Gramas+SWR	42.2	65.4	87.5	49.2
Estilística-SWR	33.8	58.9	35.6	47.3
Estilística+SWR	34.3	59.9	32.6	49.1

Fonte: Autoria própria

5.1.1 QUAL REPRESENTAÇÃO USAR?

A análise dos resultados apresentados na Tabela 5 apresenta resultados interessantes. Primeiro, a melhor abordagem para a LGTZAN é o 4-Gramas sem SWR (43.9%). Segundo, as características estilísticas obtiveram resultados similares com a abordagem *stemming* sem SWR. Terceiro, a abordagem 4-Gramas ainda provê resultados melhores do que as abordagens 2-Gramas e 3-Gramas.

A análise dos resultados da LMLD apresentados na Tabela 5 apresenta resultados similares aos analisados para a LGTZAN. A melhor abordagem na LMLD é a abordagem 4-Gramas sem remoção de *stopwords* (65.6%). Primeiro, o *stemming* obteve o pior resultado dentre todas as demais abordagens. Segundo, o 4-Gramas provê resultados melhores do que o 2-Gramas e 3-Gramas.

A melhor abordagem para a NLD (apresentada na Tabela 5) é o 4-Gramas com SWR (87.5%). Primeiro, aumentando o valor de n na representação n-gramas é possível melhorar os resultados da classificação. Segundo, a representação n-grama é melhor do que *stemming* em todas as parametrizações. Terceiro, o uso das características estilísticas provê um resultado interessante que atinge resultados de classificação similares a abordagem *stemming*, mas usando somente 17 características.

Os resultados da JLD são apresentados na Tabela 5. A melhor abordagem para a JLD é o 3-Gramas com SWR (52.2%). Este é um resultado interessante, pois para as bases de letras individuais, o 3-Gramas não atingiu os melhores resultados nas diferentes bases de dados. Para a NLD este atingiu 78.4% de *F-measure*, enquanto a melhor abordagem (4-Gramas-SWR) atingiu 87.7%. Para a LMLD este obteve 60.3% enquanto que a melhor abordagem (4-Gramas-SWR) atingiu 65.6%. Para a LGTZAN, o 3-Gramas+SWR obteve 40.9%, enquanto o melhor resultado (4-Gramas) obteve 43.9%. A análise destes resultados demonstra que enquanto alguns métodos trabalharam bem nas bases de letras individualmente, eles não trabalharam tão bem com um número maior de idiomas.

5.1.2 QUAL É O IMPACTO DAS *STOPWORDS*?

Os resultados sem e com SWR para a LGTZAN são apresentados na Tabela 5. Note que o uso de SWR claramente melhora a medida de classificação *F-measure* da abordagem *stemming* de 34.3% para 40.4%. Para as características estilísticas de texto existe um pequeno ganho em *F-measure* enquanto que para as abordagens baseadas em n-gramas sempre há perca

em *F-measure*.

Para a LMLD o uso de SWR melhora os resultados como mostrado na Tabela 5. Apenas as características estilísticas de texto são beneficiadas pelo uso da remoção das *stopwords*. Para as representações N-Gramas existe uma diminuição exceto para 3-Gramas que mantiveram o mesmo resultado. A abordagem *stemming* não obteve bons resultados com o uso de SWR diminuindo de 37.6% para 31.7%.

Ao lidar com letras em idiomas Nórdicos ainda não está claro se as *stopwords* deveriam ou não ser removidas. A análise dos resultados apresentados na Tabela 5 demonstra que removendo as *stopwords* das letras de músicas na NLD prejudica os resultados da classificação em várias representações.

Para a JLD, o uso de SWR claramente melhora os resultados para as características estilísticas de texto como apresentado na Tabela 5. Enquanto existe um pequeno ganho (0.4%) em *F-measure* para 3-Gramas e uma diminuição em *F-measure* (4.4%) para *stemming*.

5.2 AVALIAÇÃO E DISCUSSÃO GLOBAL DOS RESULTADOS

O melhor resultado obtido dentre todas as base de dados foi 87.5% para a NLD usando 4-gramas com SWR. Para as demais bases, o melhor resultado de cada uma foi: 4-Gramas-SWR para a LGTZAN e LMLD, atingindo respectivamente 43.9% e 65.6% de *F-measure*; 3-Gramas+SWR (atingindo 52.2%) para a JLD. Portanto, os resultados experimentais demonstram uma tendência para o uso da abordagem 4-gramas ao lidar com letras de músicas em diferentes idiomas.

Experimentos adicionais foram executados para avaliar o impacto da remoção das *stopwords* na tarefa de classificação de letras de músicas. Isto é importante por duas razões: não está claro se as *stopwords* deveriam ou não ser removidas ao lidar com letras de músicas; e as *stopwords* são recursos dependentes de idiomas e podem não estar disponíveis para alguns idiomas. Para as características estilísticas está claro que o uso de SWR sempre melhora os resultados nos experimentos realizados. No entanto, para as demais técnicas não houve melhoria evidente, pois em alguns casos melhorou e em outros não.

6 EXPERIMENTOS COMPUTACIONAIS

6.1 RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA LGTZAN

Nesta seção estão os resultados (em %) para as combinações das representações usando as abordagens *early* e *late fusion* (com os métodos da probabilidade máxima, soma e produto das probabilidades) para a LGTZAN apresentados na Tabela 6. Esta seção contém os resultados detalhados para a LGTZAN para as seguintes questões: A combinação com alta dimensionalidade através de N-Gramas+*Stemming* é a melhor abordagem? Ao usar as características estilísticas qual é a melhor combinação N-Gramas+Estilística, N-Gramas+*Stemming*+Estilística ou outras combinações?

Tabela 6: Resultados para os métodos *early* e *late fusion* na LGTZAN

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
2-Gramas+ <i>Stemming</i> -SWR	42.4	41.2	44.5	45.9
2-Gramas+ <i>Stemming</i> +SWR	49.7	50.2	52.2	58.4
3-Gramas+ <i>Stemming</i> -SWR	44.8	48.5	54.6	53.0
3-Gramas+ <i>Stemming</i> +SWR	52.5	60.2	59.5	63.5
4-Gramas+ <i>Stemming</i> -SWR	43.6	55.4	60.8	55.9
4-Gramas+ <i>Stemming</i> +SWR	53.4	64.6	67.6	64.7
N-Gramas+ <i>Stemming</i> -SWR	45.6	52.8	59.2	58.5
N-Gramas+ <i>Stemming</i> +SWR	43.3	52.4	60.2	59.6
2-Gramas+3-Gramas+ <i>Stemming</i> -SWR	45.1	47.1	56.2	55.4
2-Gramas+3-Gramas+ <i>Stemming</i> +SWR	52.5	56.6	57.8	58.9
2-Gramas+4-Gramas+ <i>Stemming</i> -SWR	45.4	54.2	57.6	57.1
2-Gramas+4-Gramas+ <i>Stemming</i> +SWR	52.0	58.2	62.3	64.0
3-Gramas+4-Gramas+ <i>Stemming</i> -SWR	44.4	54.4	61.3	58.6
3-Gramas+4-Gramas+ <i>Stemming</i> +SWR	53.0	64.2	61.6	62.5

Continua na próxima página

Tabela 6 – Continuação da página anterior

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
2-Gramas+Estilística-SWR	42.6	47.6	49.4	54.8
2-Gramas+Estilística+SWR	43.9	43.5	42.9	43.8
3-Gramas+Estilística-SWR	44.8	58.1	51.3	63.2
3-Gramas+Estilística+SWR	46.8	57.3	59.5	64.6
4-Gramas+Estilística-SWR	42.4	65.7	66.7	69.9
4-Gramas+Estilística+SWR	42.9	62.6	65.8	70.7
2-Gramas+3-Gramas+Estilística-SWR	50.7	53.9	61.4	78.5
2-Gramas+3-Gramas+Estilística+SWR	48.3	51.8	56.3	58.9
2-Gramas+4-Gramas+Estilística-SWR	51.9	60.2	65.6	65.2
2-Gramas+4-Gramas+Estilística+SWR	50.7	53.2	61.9	67.3
3-Gramas+4-Gramas+Estilística-SWR	50.8	64.8	65.1	67.0
3-Gramas+4-Gramas+Estilística+SWR	50.5	64.5	60.3	61.6
N-Gramas+Estilística-SWR	50.7	59.7	50.5	68.5
N-Gramas+Estilística+SWR	50.2	52.2	58.6	63.5
2-Gramas+ <i>Stemming</i> +Estilística-SWR	46.8	52.1	60.8	65.8
2-Gramas+ <i>Stemming</i> +Estilística+SWR	52.5	51.1	60.4	62.6
3-Gramas+ <i>Stemming</i> +Estilística-SWR	51.0	57.1	68.4	53.0
3-Gramas+ <i>Stemming</i> +Estilística+SWR	54.2	60.6	67.4	59.7
4-Gramas+ <i>Stemming</i> +Estilística-SWR	52.5	63.0	63.2	63.0
4-Gramas+ <i>Stemming</i> +Estilística+SWR	53.9	64.7	62.4	70.3
N-Gramas+ <i>Stemming</i> +Estilística-SWR	50.7	54.3	60.0	64.6
N-Gramas+ <i>Stemming</i> +Estilística+SWR	50.8	54.3	67.0	63.6
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística-SWR	51.2	56.6	59.5	61.2
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística+SWR	54.2	56.7	59.9	65.1
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	51.2	60.3	64.8	64.3
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	53.9	58.2	66.6	67.7
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	50.3	61.5	63.9	65.5
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	54.9	61.6	66.4	66.1
<i>Stemming</i> +Estilística-SWR	41.7	48.5	52.4	58.9
<i>Stemming</i> +Estilística+SWR	48.0	53.2	53.3	53.2
2-Gramas+3-Gramas-SWR	44.9	50.6	53.4	56.1

Continua na próxima página

Tabela 6 – *Continuação da página anterior*

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
2-Gramas+3-Gramas+SWR	40.4	49.5	48.1	49.2
2-Gramas+4-Gramas-SWR	44.9	59.1	61.3	57.8
2-Gramas+4-Gramas+SWR	43.6	49.9	51.0	55.3
3-Gramas+4-Gramas-SWR	43.2	62.0	62.4	61.0
3-Gramas+4-Gramas+SWR	42.4	61.3	62.3	61.8

Fonte: Autoria Própria

6.1.1 COMBINANDO N-GRAMAS COM REPRESENTAÇÕES *STEMMING*

Os resultados para a combinação de N-Gramas com *stemming* são apresentados na Tabela 6. Para a LGTZAN o melhor resultado usando combinação por *early fusion* (primeira coluna da Tabela 6) foi 4-Gramas+*Stemming*+SWR (53.4%). Uma observação interessante é que combinações com mais representações não necessariamente atingem os melhores resultados na LGTZAN. Desta forma, o processo de remoção de stopwords melhora a maioria dos casos, menos para N-Gramas com representações *stemming*.

Os resultados para a combinação das representações com N-Gramas e *Stemming* usando *late fusion* com o método de probabilidade máxima para LGTZAN são apresentados na segunda coluna da Tabela 6. O melhor resultado foi obtido com a combinação 4-Gramas+*Stemming*+SWR (64.6%). O uso da remoção de stopwords nesta abordagem claramente melhora os resultados atingindo ganhos de 9.8%, menos para N-Gramas+*Stemming*.

Os resultados para a combinação de N-Gramas com *stemming* para a LGTZAN usando *late fusion* com o método da soma da probabilidade são apresentados na terceira coluna da Tabela 6. O melhor resultado obtido foi 61.6% com 3-Gramas+4-Gramas+*Stemming*+SWR.

Os resultados para a combinação de N-Gramas com *stemming* usando *late fusion* com o produto da probabilidade são apresentados na quarta coluna da Tabela 6. Para a LGTZAN a combinação 4-Gramas+*Stemming*+SWR atingiu o melhor resultado (64.7%). Novamente, é importante ressaltar que as combinações com mais representações não necessariamente atingem os melhores resultados para a LGTZAN.

6.1.2 COMBINANDO N-GRAMAS COM CARACTERÍSTICAS ESTILÍSTICAS

Os resultados para a combinação das características estilísticas de texto com as abordagens n-gramas usando *early fusion* são apresentados na primeira coluna da Tabela 6. A análise demonstra resultados interessantes para a LGTZAN. Primeiro, houve uma melhora nos resultados de forma geral. Segundo, a combinação de todas as representações com as características estilísticas de texto melhorou os resultados para a maioria das abordagens, independentemente do uso ou não de SWR. Este é um resultado interessante tendo em vista que as características estilísticas de texto são um conjunto compacto de atributos. A exceção foi apenas a abordagem 4-Gramas sem SWR, na qual o melhor resultado sem combinação *no fusion* foi obtido na LGTZAN.

Os resultados para a combinação de N-Gramas e características estilísticas de texto para a LGTZAN usando *late fusion* com o método da probabilidade máxima são apresentados na segunda coluna da Tabela 6. A melhor combinação foi 4-Gramas+Estilística-SWR atingindo 65.7%. As combinações com a mais alta dimensionalidade não necessariamente atingiram os melhores resultados.

Os resultados da LGTZAN para a combinação de N-Gramas com as características estilísticas de texto usando *late fusion* com a soma das probabilidades são apresentados na terceira coluna da Tabela 6. A combinação 2-Gramas+4-Gramas+Estilística-SWR atingiu o melhor resultado na LGTZAN (67.0%). O uso do processo de remoção das stopwords foi indiferente porque em alguns casos melhorou os resultados, mas para outros não.

Os resultados da LGTZAN para a combinação N-Gramas com as características estilísticas de texto usando *late fusion* com o produto da probabilidade são apresentados na quarta coluna da Tabela 6. A combinação 4-Gramas+Estilística+SWR atingiu o melhor resultado na LGTZAN (70.7%). O uso do processo de remoção das stopwords foi indiferente, pois em alguns casos melhorou os resultados e em outros não.

6.1.3 COMBINANDO N-GRAMAS, *STEMMING* E CARACTERÍSTICAS ESTILÍSTICAS

Os resultados para a combinação de N-Gramas com *stemming* e características estilísticas usando *early fusion* são apresentados na primeira coluna da Tabela 6. Para a LGTZAN o melhor resultado obtido foi 54.9% para 3-Gramas+4-Gramas+*Stemming*+Estilística+SWR. Este resultado é melhor do que a combinação de N-Gramas com *stemming* na qual o melhor resultado foi 4-Gramas+*Stemming*+SWR (53.4%) como mostrado na Tabela 6. Além disso,

é importante ressaltar que para a LGTZAN os resultados da combinação de N-Gramas com *stemming* e Estilísticas foi melhor comparado as outras combinações.

Os resultados para a combinação de N-Gramas, *Stemming* e características estilísticas para LGTZAN usando *late fusion* com probabilidade máxima são apresentados na segunda coluna da Tabela 6. A melhor combinação foi 4-Gramas+*Stemming*+Estilística+SWR atingindo 64.7%.

Os resultados para a combinação de N-Gramas e *Stemming* com características estilísticas usando *late fusion* com a soma da probabilidade são apresentados na terceira coluna da Tabela 6. A melhor combinação foi 4-Gramas+*Stemming*+Estilística-SWR atingindo 68.4%.

Os resultados para a combinação de N-Gramas com *Stemming* e características estilísticas usando *late fusion* com o produto da probabilidade para LGTZAN são apresentados na quarta coluna da Tabela 6. A melhor combinação foi 4-Gramas+*Stemming*+Estilística+SWR atingindo 70.3%.

6.1.4 OUTRAS COMBINAÇÕES

Os resultados para outras combinações usando a abordagem *early fusion* são apresentados na primeira coluna da Tabela 6. Para a LGTZAN, a abordagem *stemming* com SWR combinada com as características estilísticas obteve uma *F-measure* de 48.0% (contra 40.4% sem uso de características estilísticas). Este resultado é um bom valor para a LGTZAN o qual melhora o resultado *stemming no fusion*.

Os resultados da LGTZAN para outras combinações usando *late fusion* com probabilidade máxima são apresentados na segunda coluna da Tabela 6. O melhor resultado nesta análise específica foi 3-Gramas+4-Gramas-SWR atingindo 62.0%.

Os resultados da LGTZAN para outras combinações usando *late fusion* com soma da probabilidade são apresentados na terceira coluna da Tabela 6. A melhor representação foi 3-Gramas+4-Gramas-SWR atingindo 62.4%. Note que este resultado não foi satisfatório comparando com outras combinações.

Os resultados da LGTZAN para outras combinações usando *late fusion* com o produto da probabilidade são apresentados na quarta coluna da Tabela 6. O melhor resultado nesta análise específica foi 3-Gramas+4-Gramas+SWR atingindo 61.8%.

6.2 RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA LMLD

Nesta seção estão os resultados (em %) para as combinações das representações para a LMLD apresentados na Tabela 7. Esta seção contém os resultados detalhados para as seguintes questões: A abordagem de combinação com maior dimensionalidade com N-Gramas+*Stemming* é a melhor abordagem? A utilização das características estilísticas é melhor com a combinação N-Gramas+Estilística, N-Gramas+*Stemming*+Estilística ou outras combinações?

Tabela 7: Resultados para os métodos *early* e *late fusion* na LMLD

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
2-Gramas+ <i>Stemming</i> -SWR	53.8	55.9	57.3	57.3
2-Gramas+ <i>Stemming</i> +SWR	53.6	56.1	58.1	57.9
3-Gramas+ <i>Stemming</i> -SWR	60.2	64.1	65.0	64.1
3-Gramas+ <i>Stemming</i> +SWR	61.8	64.4	66.1	66.7
4-Gramas+ <i>Stemming</i> -SWR	65.8	71.4	71.6	69.5
4-Gramas+ <i>Stemming</i> +SWR	66.2	71.2	71.1	63.8
N-Gramas+ <i>Stemming</i> -SWR	67.4	72.1	73.9	62.0
N-Gramas+ <i>Stemming</i> +SWR	67.4	72.2	74.3	49.9
2-Gramas+3-Gramas+ <i>Stemming</i> -SWR	61.0	65.1	66.3	66.0
2-Gramas+3-Gramas+ <i>Stemming</i> +SWR	63.3	67.4	68.7	69.5
2-Gramas+4-Gramas+ <i>Stemming</i> -SWR	66.4	71.3	72.7	65.0
2-Gramas+4-Gramas+ <i>Stemming</i> +SWR	66.3	71.4	73.3	54.4
3-Gramas+4-Gramas+ <i>Stemming</i> -SWR	67.3	72.6	72.8	66.6
3-Gramas+4-Gramas+ <i>Stemming</i> +SWR	67.2	72.2	72.9	58.0
2-Gramas+Estilística-SWR	73.5	70.0	76.3	80.7
2-Gramas+Estilística+SWR	71.8	69.0	75.2	77.7
3-Gramas+Estilística-SWR	74.8	72.6	74.8	79.4
3-Gramas+Estilística+SWR	75.5	74.2	76.7	79.1
4-Gramas+Estilística-SWR	77.3	73.6	75.3	78.4
4-Gramas+Estilística+SWR	77.5	75.0	75.8	75.7
2-Gramas+3-Gramas+Estilística-SWR	75.8	71.9	76.3	80.2
2-Gramas+3-Gramas+Estilística+SWR	76.5	73.7	77.3	67.7
2-Gramas+4-Gramas+Estilística-SWR	78.2	71.9	77.6	70.3
2-Gramas+4-Gramas+Estilística+SWR	79.3	73.8	77.9	68.5

Continua na próxima página

Tabela 7 – Continuação da página anterior

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
3-Gramas+4-Gramas+Estilística-SWR	78.3	72.7	76.9	70.6
3-Gramas+4-Gramas+Estilística+SWR	79.1	73.5	78.4	80.2
N-Grams+Estilística-SWR	78.4	73.6	78.3	80.5
N-Grams+Estilística+SWR	78.7	73.3	79.5	80.1
2-Gramas+ <i>Stemming</i> +Estilística-SWR	74.9	69.8	73.7	79.0
2-Gramas+ <i>Stemming</i> +Estilística+SWR	74.5	68.8	75.1	77.8
3-Gramas+ <i>Stemming</i> +Estilística-SWR	76.0	71.9	75.5	78.9
3-Gramas+ <i>Stemming</i> +Estilística+SWR	76.1	73.6	77.3	80.2
4-Gramas+ <i>Stemming</i> +Estilística-SWR	76.9	72.5	76.7	63.9
4-Gramas+ <i>Stemming</i> +Estilística+SWR	79.1	74.4	77.8	78.2
N-Gramas+ <i>Stemming</i> +Estilística-SWR	79.0	73.3	74.3	57.6
N-Gramas+ <i>Stemming</i> +Estilística+SWR	79.3	74.3	79.8	44.5
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística-SWR	76.0	71.6	76.7	80.6
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística+SWR	76.7	73.7	78.5	80.9
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	78.3	71.7	79.0	76.7
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	79.6	74.0	78.9	74.0
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	78.6	72.5	77.6	76.4
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	79.2	73.4	79.1	73.3
<i>Stemming</i> +Estilística-SWR	65.6	62.6	70.2	73.8
<i>Stemming</i> +Estilística+SWR	64.6	65.3	70.2	73.8
2-Gramas+3-Gramas-SWR	60.6	65.3	65.2	67.3
2-Gramas+3-Gramas+SWR	62.7	67.4	67.7	69.3
2-Gramas+4-Gramas-SWR	66.4	71.1	71.6	71.5
2-Gramas+4-Gramas+SWR	65.9	71.6	71.3	71.2
3-Gramas+4-Gramas-SWR	67.2	72.7	72.5	72.5
3-Gramas+4-Gramas+SWR	66.6	72.3	72.0	72.2

Fonte: Autoria própria

6.2.1 COMBINANDO REPRESENTAÇÕES N-GRAMAS COM *STEMMING*

A combinação 3-Gramas+4-Gramas+*Stemming*-SWR usando *early fusion* atingiu 72.6% de *F-measure* na LMLD como apresentado na primeira coluna da Tabela 7. Em geral, as

combinações com a dimensionalidade mais alta atingiram os melhores resultados ao combinar N-Gramas e *Stemming*. O processo de remoção de stopwords não produziu melhorias claras nos resultados da LMLD.

O melhor resultado atingido usando *late fusion* com o método da probabilidade máxima na LMLD foi 67.4% para a combinação N-Gramas com *stemming* como apresentado na segunda coluna da Tabela 7. Note que os resultados da LMLD diferem da LGTZAN na qual os melhores resultados são obtidos pelas combinações com a dimensionalidade mais alta. Para a LMLD o processo de remoção de stopwords não demonstrou melhorias claras exceto para 3-Gramas+*Stemming*, 4-Gramas+*Stemming* e 2-Gramas+3-Gramas+*Stemming* nas quais os resultados melhoraram em até 2.3%.

A melhor combinação da LMLD usando *late fusion* com a soma da probabilidade (terceira coluna como apresentado na Tabela 7) foi N-Gramas+*Stemming*+SWR atingindo 74.3%. Em geral o processo de remoção das stopwords melhorou os resultados exceto para a combinação 4-Gramas+*Stemming*.

O melhor resultado atingido na LMLD foi 69.5% para as combinações 2-Gramas+3-Gramas+*Stemming*+SWR e 4-Gramas+*Stemming*-SWR usando *late fusion* com o método do produto da probabilidade como mostrado na quarta coluna da Tabela 7.

6.2.2 COMBINANDO N-GRAMAS COM CARACTERÍSTICAS ESTILÍSTICAS

Para a LMLD usando *early fusion* (primeira coluna como apresentado na Tabela 7) a melhor combinação 4-Gramas+Estilística+SWR atingiu 75.0%. As outras combinações retornaram resultados similares nos quais o pior resultado foi 69.0% para a combinação 2-Gramas+Estilística+SWR.

A análise dos resultados apresentados na segunda coluna da Tabela 7 indica que para a LMLD a combinação das abordagens sem combinação (*no fusion*) com as características estilísticas usando *late fusion* com o método da probabilidade máxima melhora a classificação *F-measure* na maioria dos casos, provendo melhores resultados do que as representações sem combinação (*no fusion*).

Para a LMLD a melhor combinação usando *late fusion* com o método da soma da probabilidade (terceira coluna da Tabela 7) foi N-Gramas+Estilística+SWR atingindo 79.5%. É importante ressaltar que diferentemente da abordagem *late fusion* com o método da probabilidade máxima os resultados melhoraram com as combinações de dimensionalidade mais alta

para a LMLD.

A melhor combinação da LMLD usando *late fusion* com o produto da probabilidade (quarta coluna da Tabela 7) foi 2-Gramas+Estilística-SWR atingindo 80.7%. Perceba que esta combinação tem poucas características e atingiu o segundo melhor resultado na LMLD.

6.2.3 COMBINANDO N-GRAMAS, *STEMMING* E CARACTERÍSTICAS ESTILÍSTICAS

Para a LMLD a melhor combinação usando *early fusion* (primeira coluna da Tabela 7) foi a mesma da LGTZAN (4-Gramas+*Stemming*+Estilística+SWR) atingindo 74.4%. Note que novamente a combinação de N-Gramas com *Stemming* e características estilísticas não foi melhor que N-Gramas e características estilísticas apenas.

A análise dos resultados apresentados na segunda coluna da Tabela 7 demonstra que para a LMLD a combinação usando *late fusion* com probabilidade máxima de N-Gramas com *stemming* e características estilísticas também melhora os resultados comparando a outras abordagens. A combinação 2-Gramas+4-Gramas+*Stemming*+Estilística+SWR atingiu o melhor resultado (79.6%). Perceba que a performance diferente entre as combinações foi pequena na qual o pior resultado obtido foi 74.5% para 2-Gramas+*Stemming*+Estilística+SWR (a combinação com o espaço de dimensionalidade menor).

Para a LMLD a melhor combinação usando *late fusion* com a soma da probabilidade (terceira coluna da Tabela 7) foi N-Gramas+*Stemming*+Estilística+SWR atingindo 79.8%. Este resultado é melhor do que outras abordagens previamente analisadas.

Para a LMLD a melhor combinação usando *late fusion* com o produto da probabilidade (quarta coluna da Tabela 7) foi 2-Gramas+3-Gramas+*Stemming*+Estilística+SWR atingindo 80.9%.

6.2.4 OUTRAS COMBINAÇÕES

Para a LMLD a melhor combinação usando *early fusion* (primeira coluna da Tabela 7) foi o mesmo da LGTZAN (3-Gramas+4-Gramas-SWR) atingindo 72.7%. Nas letras de músicas da LMLD também houve melhorias dos resultados sem combinação (*no fusion*) para as combinações usando *early fusion*. Por exemplo, o melhor resultado foi 3-Gramas+4-Gramas-SWR (67.2%) usando *early fusion*, exceto outras combinações que superaram este resultado.

Para a LMLD a melhor combinação foi 3-Gramas+4-Gramas-SWR atingindo 72.5%.

Perceba que esta combinação não foi a melhor ao comparar as abordagens usadas com *late fusion* com a soma das probabilidades. Mas, o melhor resultado geral para a LMLD foi *Stemming*+Estilística atingindo 73.8% usando *late fusion* com o produto da probabilidade.

6.3 RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA NLD

Nesta seção a análise dos resultados (em %) tem foco nas combinações usando as abordagens *early fusion* e *late fusion* apresentadas na Tabela 8. Esta seção contém os resultados detalhados para as seguintes questões: a combinação com a dimensionalidade mais alta com N-Gramas+*Stemming* é a melhor abordagem? Ao usar características estilísticas é melhor combinar com N-Gramas+Estilística, N-Gramas+*Stemming*+Estilística ou outra combinação?

Tabela 8: Resultados para os métodos *early* e *late fusion* na NLD

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
2-Gramas+ <i>Stemming</i> -SWR	62.0	64.5	64.3	65.5
2-Gramas+ <i>Stemming</i> +SWR	62.6	65.9	64.9	64.4
3-Gramas+ <i>Stemming</i> -SWR	77.9	81.5	80.8	81.1
3-Gramas+ <i>Stemming</i> +SWR	78.6	81.5	81.3	81.0
4-Gramas+ <i>Stemming</i> -SWR	87.4	80.3	80.6	71.9
4-Gramas+ <i>Stemming</i> +SWR	87.7	81.0	82.1	75.7
N-Gramas+ <i>Stemming</i> -SWR	94.7	85.8	85.2	77.7
N-Gramas+ <i>Stemming</i> +SWR	95.5	87.6	88.4	81.9
2-Gramas+3-Gramas+ <i>Stemming</i> -SWR	82.4	81.5	80.5	83.7
2-Gramas+3-Gramas+ <i>Stemming</i> +SWR	82.7	82.5	81.1	83.6
2-Gramas+4-Gramas+ <i>Stemming</i> -SWR	88.3	80.5	79.8	71.0
2-Gramas+4-Gramas+ <i>Stemming</i> +SWR	88.3	80.5	81.7	74.6
3-Gramas+4-Gramas+ <i>Stemming</i> -SWR	94.4	86.5	86.5	81.3
3-Gramas+4-Gramas+ <i>Stemming</i> +SWR	94.3	88.4	87.7	83.8
2-Gramas+Estilística-SWR	64.2	64.7	66.8	68.5
2-Gramas+Estilística+SWR	64.2	64.4	66.9	68.6
3-Gramas+Estilística-SWR	78.1	80.9	81.1	81.4
3-Gramas+Estilística+SWR	78.4	80.9	81.6	82.2
4-Gramas+Estilística-SWR	87.9	80.1	81.6	83.0
4-Gramas+Estilística+SWR	87.4	81.9	82.4	84.6

Continua na próxima página

Tabela 8 – Continuação da página anterior

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
2-Gramas+3-Gramas+Estilística-SWR	83.5	79.9	84.5	85.4
2-Gramas+3-Gramas+Estilística+SWR	84.0	81.0	84.2	85.1
2-Gramas+4-Gramas+Estilística-SWR	89.2	79.3	82.9	65.1
2-Gramas+4-Gramas+Estilística+SWR	89.4	79.1	84.5	71.6
3-Gramas+4-Gramas+Estilística-SWR	94.2	84.6	90.0	74.7
3-Gramas+4-Gramas+Estilística+SWR	94.5	86.8	88.8	80.9
N-Gramas+Estilística-SWR	94.9	85.5	88.4	93.1
N-Gramas+Estilística+SWR	94.3	87.3	88.8	93.1
2-Gramas+ <i>Stemming</i> +Estilística-SWR	64.2	62.7	66.3	67.6
2-Gramas+ <i>Stemming</i> +Estilística+SWR	64.1	63.5	65.9	67.7
3-Gramas+ <i>Stemming</i> +Estilística-SWR	79.8	80.0	82.1	81.9
3-Gramas+ <i>Stemming</i> +Estilística+SWR	80.0	80.3	82.2	82.0
4-Gramas+ <i>Stemming</i> +Estilística-SWR	88.1	78.0	81.7	73.3
4-Gramas+ <i>Stemming</i> +Estilística+SWR	88.7	79.2	83.7	76.7
N-Gramas+ <i>Stemming</i> +Estilística-SWR	95.0	85.5	86.6	73.9
N-Gramas+ <i>Stemming</i> +Estilística+SWR	95.5	87.5	88.6	78.5
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística-SWR	83.4	79.9	81.5	83.7
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística+SWR	83.0	81.0	81.1	83.8
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	89.1	79.2	81.1	71.5
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	89.5	78.8	82.1	74.8
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	94.5	84.6	88.4	81.0
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	94.8	86.1	89.1	83.2
<i>Stemming</i> +Estilística-SWR	47.0	44.1	48.5	47.0
<i>Stemming</i> +Estilística+SWR	43.6	41.7	44.9	43.7
2-Gramas+3-Gramas-SWR	82.3	81.7	82.8	84.1
2-Gramas+3-Gramas+SWR	82.7	82.3	83.0	84.6
2-Gramas+4-Gramas-SWR	88.7	80.7	80.3	82.9
2-Gramas+4-Gramas+SWR	88.5	80.6	81.4	82.9
3-Gramas+4-Gramas-SWR	94.4	86.4	88.3	92.0
3-Gramas+4-Gramas+SWR	94.4	88.3	89.4	93.1

Fonte: Autoria própria

6.3.1 COMBINANDO AS REPRESENTAÇÕES N-GRAMAS COM *STEMMING*

Os resultados para a combinação das representações N-Gramas e *Stemming* para a NLD usando *late fusion* com probabilidade máxima são mostrados na segunda coluna da Tabela 8. A melhor combinação (3-Gramas+4-Gramas+*Stemming*+SWR) resultou em 88.4% de *F-measure*. Adicionalmente, note que em geral houve melhorias para os resultados sem combinação (*no fusion*). Estas melhorias não justificam a combinação das representações usando *late fusion* com o método da probabilidade máxima o qual impacta na complexidade da classificação.

Os resultados para a NLD usando *late fusion* com a soma da probabilidade seguem os padrões da LMLD na qual representações com alta dimensionalidade apresentaram os melhores resultados como mostrado na Tabela 8. O melhor resultado obtido para a NLD foi 95.5% para a combinação N-Gramas+*Stemming*+SWR. Entretanto, o processo de remoção das stopwords para as letras da NLD claramente melhorou os resultados exceto a combinação 3-Gramas+4-Gramas+*Stemming*.

Os resultados para a NLD usando *late fusion* com o produto da probabilidade são apresentados na quarta coluna da Tabela 8. O melhor resultado obtido para a NLD foi 83.8% para a combinação 3-Gramas+4-Gramas+*Stemming*+SWR.

6.3.2 COMBINANDO N-GRAMAS COM CARACTERÍSTICAS ESTILÍSTICAS

Os resultados da NLD para a combinação de N-Gramas com características estilísticas usando *early fusion* são apresentados na primeira coluna da Tabela 8. A combinação 3-Gramas+4-Gramas+Estilística-SWR atingiu o melhor resultado para a NLD (90.0%). Este resultado da combinação superou todas as abordagens sem combinação (*no fusion*).

Os resultados para a combinação das representações N-Gramas e características estilísticas usando *late fusion* com probabilidade máxima para a NLD são apresentados na segunda coluna da Tabela 8. A melhor combinação foi N-Gramas+Estilística+SWR atingindo 87.3%. Entretanto, este resultado não superou o melhor resultado sem combinação (*no fusion*) (87.5%) como mostrado na Tabela 5.

A combinação das características estilísticas com as representações N-Gramas demonstrou um pequeno ganho em *F-measure* para a NLD usando *late fusion* com a soma da probabilidade (terceira coluna da Tabela 8). Mais precisamente de 62.9% para 64.2% no 2-

Gramas; de 77.4% para 78.1% em 3-Gramas e de 87.1% para 87.9% em 4-Gramas.

Os resultados da NLD para a combinação de N-Gramas com características estilísticas usando *late fusion* com o produto da probabilidade são apresentados na quarta coluna da Tabela 8. A combinação N-Gramas+Estilística atingiu o melhor resultado para NLD (93.1%). Este resultado da combinação superou as abordagens sem combinação (*no fusion*) e outras *late fusion*.

6.3.3 COMBINANDO N-GRAMAS, *STEMMING* E CARACTERÍSTICAS ESTILÍSTICAS

Os resultados da NLD para a combinação de N-Gramas com *Stemming* e características estilísticas usando *early fusion* são apresentados na primeira coluna da Tabela 8. A melhor combinação foi 3-Gramas+4-Gramas+*Stemming*+Estilística+SWR atingindo 89.1%.

Os resultados para a combinação de N-Gramas com *Stemming* e características estilísticas usando *late fusion* com probabilidade máxima para a NLD são apresentados na segunda coluna da Tabela 8. A melhor combinação foi N-Gramas+*Stemming*+Estilística+SWR atingindo 87.5%.

A análise dos resultados apresentados na Tabela 8 demonstra que para a NLD a combinação N-Gramas+*Stemming*+Estilística+SWR atingiu o melhor resultado (95.5%). Para a NLD claramente as combinações com a mais alta dimensionalidade obtém os melhores resultados. Entretanto, é importante ressaltar que a adição das características estilísticas não alterou o melhor resultado obtido usando apenas N-Grams+*Stemming*+SWR.

Os resultados para a combinação de N-Gramas com *Stemming* e características estilísticas usando *late fusion* com o produto da probabilidade para a NLD são apresentados na quarta coluna da Tabela 8. A melhor combinação foi 2-Gramas+3-Gramas+*Stemming*+Estilística+SWR atingindo 83.8%.

6.3.4 OUTRAS COMBINAÇÕES

Os resultados da NLD para as demais combinações usando *early fusion* são apresentados na primeira coluna da Tabela 8. A melhor combinação foi 3-Gramas+4-Gramas+SWR a qual atingiu 89.4% de *F-Measure*.

Os resultados das demais combinações usando *late fusion* com probabilidade máxima para a NLD são apresentados na Tabela 8. A melhor combinação foi 3-Gramas+4-Gramas+SWR

atingindo 88.3% de *F-measure*. Este resultado superou as demais combinações (e abordagens sem combinação (*no fusion*) analisadas usando a abordagem *late fusion* com o método da probabilidade máxima.

Os resultados das demais combinações usando *late fusion* com soma da probabilidade para a NLD são apresentados na terceira coluna da Tabela 8. A melhor combinação foi 3-Gramas+4-Gramas+SWR atingindo 93.1% de *F-measure*. Este resultado superou as demais combinações (e abordagens *no fusion*) analisadas usando a abordagem *late fusion* com o método da probabilidade máxima.

6.4 RESULTADOS PARA A COMBINAÇÃO DE CLASSIFICADORES NA JLD

Nesta seção a análise dos resultados (em %) para as combinações usando as abordagens *early* e *late fusion* são apresentados na Tabela 9. Esta seção contém os resultados detalhados para as seguintes questões: A combinação com a dimensionalidade mais alta (N-Gramas+*Stemming*) é a melhor? O uso das características estilísticas com a combinação N-Gramas+Estilística, N-Gramas+*Stemming*+Estilística ou outras combinações? Além disso, o que acontece com todas essas questões quando os três conjuntos de dados são combinados na JLD?

Tabela 9: Resultados para os métodos *early* e *late fusion* na JLD

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
2-Gramas+ <i>Stemming</i> -SWR	44.7	47.2	46.8	45.8
2-Gramas+ <i>Stemming</i> +SWR	44.6	47.1	47.2	46.9
3-Gramas+ <i>Stemming</i> -SWR	44.6	53.1	53.6	54.1
3-Gramas+ <i>Stemming</i> +SWR	45.1	53.2	53.9	55.8
4-Gramas+ <i>Stemming</i> -SWR	60.6	50.6	53.8	52.9
4-Gramas+ <i>Stemming</i> +SWR	61.1	51.5	52.8	54.3
N-Gramas+ <i>Stemming</i> -SWR	65.6	58.4	62.0	60.7
N-Gramas+ <i>Stemming</i> +SWR	66.7	54.9	60.6	59.9
2-Gramas+3-Gramas+ <i>Stemming</i> -SWR	53.4	55.6	58.5	56.7
2-Gramas+3-Gramas+ <i>Stemming</i> +SWR	53.2	56.0	59.0	59.9
2-Gramas+4-Gramas+ <i>Stemming</i> -SWR	50.2	55.7	59.0	58.8
2-Gramas+4-Gramas+ <i>Stemming</i> +SWR	50.5	54.2	57.3	56.7

Continua na próxima página

Tabela 9 – Continuação da página anterior

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
3-Gramas+4-Gramas+ <i>Stemming</i> -SWR	57.6	56.4	58.1	57.9
3-Gramas+4-Gramas+ <i>Stemming</i> +SWR	57.8	53.7	56.1	55.3
2-Gramas+Estilística-SWR	44.0	55.3	64.3	71.4
2-Gramas+Estilística+SWR	41.9	54.7	64.5	69.4
3-Gramas+Estilística-SWR	45.7	60.1	63.2	69.4
3-Gramas+Estilística+SWR	46.5	59.2	63.9	68.9
4-Gramas+Estilística-SWR	59.4	55.4	59.9	65.4
4-Gramas+Estilística+SWR	59.0	55.4	58.9	64.3
2-Gramas+3-Gramas+Estilística-SWR	65.6	53.9	65.7	71.5
2-Gramas+3-Gramas+Estilística+SWR	64.9	54.5	66.6	70.8
2-Gramas+4-Gramas+Estilística-SWR	61.3	54.8	65.4	70.0
2-Gramas+4-Gramas+Estilística+SWR	61.0	54.1	64.4	67.8
3-Gramas+4-Gramas+Estilística-SWR	52.5	52.9	62.8	67.2
3-Gramas+4-Gramas+Estilística+SWR	52.3	51.7	62.0	67.7
N-Gramas+Estilística-SWR	61.2	60.8	66.3	69.3
N-Gramas+Estilística+SWR	66.0	58.7	65.1	67.6
2-Gramas+ <i>Stemming</i> +Estilística-SWR	43.9	52.6	61.5	68.4
2-Gramas+ <i>Stemming</i> +Estilística+SWR	44.5	55.1	63.1	65.8
3-Gramas+ <i>Stemming</i> +Estilística-SWR	45.8	53.0	64.1	69.4
3-Gramas+ <i>Stemming</i> +Estilística+SWR	45.4	54.0	65.9	69.4
4-Gramas+ <i>Stemming</i> +Estilística-SWR	60.0	53.9	62.9	66.5
4-Gramas+ <i>Stemming</i> +Estilística+SWR	60.2	52.2	61.7	66.2
N-Gramas+ <i>Stemming</i> +Estilística-SWR	64.1	60.4	67.9	70.0
N-Gramas+ <i>Stemming</i> +Estilística+SWR	64.5	57.1	66.4	57.1
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística-SWR	65.6	54.6	66.6	69.1
2-Gramas+3-Gramas+ <i>Stemming</i> +Estilística+SWR	65.1	54.8	67.7	71.2
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	62.3	54.0	65.4	69.6
2-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	65.4	53.0	66.2	68.1
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística-SWR	58.2	52.2	63.8	66.4
3-Gramas+4-Gramas+ <i>Stemming</i> +Estilística+SWR	58.6	41.8	63.1	67.4
<i>Stemming</i> +Estilística-SWR	39.2	53.6	56.3	61.0

Continua na próxima página

Tabela 9 – Continuação da página anterior

Combinação	<i>Early Fusion</i>	Max	Soma	Produto
<i>Stemming</i> +Estilística+SWR	39.0	43.0	56.3	60.5
2-Gramas+3-Gramas-SWR	53.3	57.4	56.7	59.2
2-Gramas+3-Gramas+SWR	52.3	56.4	57.8	57.1
2-Gramas+4-Gramas-SWR	50.3	55.7	56.6	55.6
2-Gramas+4-Gramas+SWR	49.8	54.6	55.0	57.1
3-Gramas+4-Gramas-SWR	57.4	56.4	56.5	56.9
3-Gramas+4-Gramas+SWR	57.5	53.2	53.5	53.9

Fonte: Autoria própria

6.4.1 COMBINANDO N-GRAMAS COM REPRESENTAÇÕES *STEMMING*

Os resultados para a combinação de N-Gramas com *stemming* usando *early fusion* para a JLD são apresentados na primeira coluna da Tabela 9. O melhor resultado obtido foi 60.7% para a combinação N-Gramas+*Stemming*-SWR.

Na JLD o melhor resultado foi 62.0% com a combinação N-Gramas+*Stemming*-SWR usando *late fusion* com probabilidade máxima como apresentado na segunda coluna da Tabela 9.

No caso da JLD o melhor resultado obtido foi 58.4% com a combinação N-Gramas +*Stemming*-SWR como mostrado na Tabela 9. É importante ressaltar que a abordagem *late fusion* com o método da probabilidade máxima melhorou os resultados sem combinação (*no fusion*) para a JLD.

Os resultados para a combinação de N-Gramas com *stemming* usando *late fusion* com a soma da probabilidade para a JLD são apresentados na terceira coluna da Tabela 9. O melhor resultado obtido foi 66.7% para a combinação N-Gramas+*Stemming*+SWR. Note que este resultado foi similar para a LMLD e NLD demonstrando que a melhor abordagem é usar todas as representações combinadas. Para a JLD o processo de remoção das stopwords atingiu pequenas melhorias na maioria dos casos exceto 2-Gramas+*Stemming* e 2-Gramas+3-Gramas+*Stemming*.

É importante ressaltar que a combinação de N-Gramas com *stemming* melhorou os resultados sem combinação (*no fusion*) (apresentados na Tabela 5). Isto demonstra que as representações combinadas podem melhorar a classificação de gêneros musicais.

6.4.2 COMBINANDO N-GRAMAS COM CARACTERÍSTICAS ESTILÍSTICAS

Na JLD a melhor combinação usando *early fusion* (primeira coluna da Tabela 9) foi 2-Gramas+3-Gramas+Estilística-SWR atingindo 71.5%.

Na JLD a melhor combinação foi 2-Gramas+3-Gramas+Estilística+SWR atingindo 66.6% usando *late fusion* com probabilidade máxima (segunda coluna da Tabela 9). Note que todos os resultados nas outras bases de letras de músicas são diferentes deste. Isto demonstra que as combinações alteram os resultados entre as bases de dados e que nenhuma abordagem é indicada em todos os casos na combinação N-Gramas e características estilísticas.

Analisando os resultados para a JLD de acordo com a Tabela 9 note que as combinações com alta dimensionalidade não necessariamente retornam os melhores resultados exceto N-Gramas+Estilística-SWR que atingiu 60.8% usando *late fusion* com a soma da probabilidade. Em geral, o processo de remoção das stopwords foi confuso porque em alguns casos apresentou melhoras, mas não em todos os casos.

Os resultados para a combinação das características estilísticas com abordagens N-Gramas usando *late fusion* com o produto da probabilidade para a JLD são apresentados na quarta coluna da Tabela 9. A análise dos resultados demonstra que a combinação de características estilísticas de texto com outras abordagens melhora os resultados da classificação para todas as abordagens. Além disso, o melhor resultado sem combinação (*no fusion*) através da combinação de características estilísticas com N-Gramas e com SWR obteve o melhor resultado geral para a JLD (66.0%).

6.4.3 COMBINANDO N-GRAMAS, *STEMMING* E CARACTERÍSTICAS ESTILÍSTICAS

Analisando os resultados para a JLD (abordagem *early fusion*) de acordo com a Tabela 9 note que a combinação 2-Gramas+3-Gramas+*Stemming*+Estilística+SWR atingiu o melhor resultado (71.2%).

Na JLD a melhor combinação foi N-Gramas+*Stemming*+Estilística-SWR atingindo 67.9% usando *late fusion* com probabilidade máxima. Este resultado é melhor do que outras abordagens além de obter os melhores valores de *F-measure* comparando a outras abordagens *late fusion*.

Analisando os resultados para a JLD de acordo com a Tabela 9 note que a combinação com a mais alta dimensionalidade (N-Gramas+*Stemming*+Estilística-SWR) atingiu o melhor

resultado (60.4%) usando *late fusion* com soma da probabilidade. Em geral, o processo de remoção das stopwords foi confuso porque em alguns casos apresentou melhorias, mas não em todos os casos. Novamente houve algumas percas com o uso do processo de remoção das stopwords.

A combinação 2-Gramas+3-Gramas+*Stemming*+Estilística-SWR atingiu o melhor resultado (65.6%) para JLD usando *late fusion* com o produto da probabilidade como mostrado na Tabela 9. Note que este resultado não superou N-Gramas+*Stemming*+SWR (66.7%). Uma observação importante refere a diferença das melhores combinações entre as bases de letras de músicas demonstrando que não existe um padrão que indica a melhor abordagem para a classificação de gêneros musicais utilizando letras de músicas com *early fusion*.

6.4.4 OUTRAS COMBINAÇÕES

Analisando os resultados para a JLD de acordo com a Tabela 9 note que a melhor combinação foi *Stemming*+Estilística-SWR atingindo 61.0% usando o método *early fusion*.

Na JLD a melhor combinação foi 2-Gramas+3-Gramas+SWR atingindo 57.8% usando *late fusion* com probabilidade máxima.

Analisando os resultados para a JLD de acordo com a Tabela 9 note que houve empate para as combinações 2-Gramas+3-Gramas+SWR e 3-Gramas+4-Gramas-SWR (ambos atingiram 56.4%). O processo de remoção das stopwords diminuiu todos os resultados analisados nesta seção.

A análise dos resultados apresentados na Tabela 9 demonstra que a combinação das características estilísticas com a abordagem *stemming* melhorou os resultados da classificação de 38.7% (usando somente *stemming*) para 47%. Entretanto, a melhor combinação é 3-Gramas com 4-Gramas que atingiu 94.4%.

A análise dos resultados da JLD apresentados na Tabela 9 demonstra que houve pequenas melhorias. Por exemplo, combinando 3-Gramas com 4-Gramas usando SWR resultou em 57.5% a qual é melhor do que outros resultados sem combinação (*no fusion*).

6.5 AVALIAÇÃO E DISCUSSÃO GLOBAL DOS RESULTADOS

A avaliação resumida dos resultados é apresentada na Seção 6.5.1. Para avaliar se existe diferença estatística entre os métodos utilizados foram aplicados os testes Friedman e

Nemenyi indicados pelo trabalho de (DEMŠAR, 2006) na Seção 6.5.2. Na Seção 6.5.3 é apresentada a análise dos testes estatísticos executados sobre os resultados obtidos.

6.5.1 AVALIAÇÃO DOS RESULTADOS

Analisando todos os resultados das quatro bases de dados de letras de músicas utilizadas neste trabalho verificou-se que as melhores abordagens globais foram N-Gramas+*Stemming*+SWR e N-Gramas+*Stemming*+Estilística+SWR usando *early fusion* na NLD, em que ambas atingiram 95.5% de *F-measure*. Além disso, nas bases de dados LGTZAN e JLD, por exemplo, o melhor resultado foi atingido com a combinação 2-Gramas+3-Gramas+Estilística-SWR (78.5% para a LGTZAN e 71.5% para a JLD) usando *late fusion* com o produto das probabilidades, sendo esta uma representação com dimensionalidade alta.

Quanto a análise da combinação das demais características com *Stemming* para verificar se esta é a melhor escolha, foi observado que o uso de *Stemming* foi vantajoso na NLD com N-Gramas+*Stemming*+SWR e N-Gramas+*Stemming*+Estilística+SWR usando *early fusion* (atingindo 95.5% de *F-measure*) e na LMLD com a representação 2-Gramas+3-Gramas+*Stemming*+Estilística-SWR (atingindo 80.9% de *F-measure*) usando *late fusion* com produto das probabilidades.

Outra verificação realizada foi o benefício ou não de combinar as características estilísticas de texto com outras abordagens utilizadas neste trabalho. A análise dos resultados sugere que estas devem ser utilizadas com outras representações, pois comumente melhoram os resultados da classificação. Há somente uma exceção para não combinar quando as características estilísticas de texto não são capazes de classificar corretamente os gêneros musicais quando utilizadas de forma individual. Além disso, foi verificado que as características estilísticas obtêm bons resultados na combinação com todas as representações N-Gramas utilizando *late fusion* com produto da probabilidade. Entre os resultados obtidos estão: 68.5% na NLD e 80.5% na LMLD com a combinação N-Gramas+Estilística-SWR; 93.1% na LGTZAN utilizando as combinações N-Gramas+Estilística-SWR e N-Gramas+Estilística+SWR; e 69.3% na JLD com a combinação N-Gramas+Estilística-SWR.

6.5.2 TESTE FRIEDMAN E NEMENYI

O teste Friedman (FRIEDMAN, 1937, 1940) é não paramétrico. Este teste ranqueia os algoritmos para cada conjunto de dados separadamente, em que o melhor algoritmo fica na

posição 1, o segundo melhor na posição 2, e assim por diante.

Se a hipótese nula é rejeitada, então é possível dar procedência com um teste *post-hoc*. O teste Nemenyi (NEMENYI, 1962) é usado quando todos os classificadores são comparados com os demais. A performance de dois classificadores é significativamente diferente se a média da posição do algoritmo correspondente difere pelo menos da diferença crítica

$$CD = q_{\alpha} \sqrt{\frac{k(k+1)}{6N}} \quad (8)$$

onde q_{α} são valores críticos.

6.5.3 ANÁLISE DOS TESTES ESTATÍSTICOS

O *ranking* obtido pelos testes estatísticos é apresentado na Tabela 10. Considerando $q_{\alpha} = 0.05$ obtém-se uma diferença crítica de 198.218, desta forma os métodos que possuem diferença estatística são: PROD+N-Grams+Est-SWR (9.25), Stem-SWR (208.375) e Stem+SWR (208.375). Em contra partida, ao considerar $q_{\alpha} = 0.10$ obtém-se uma diferença crítica de 190.420, assim, os métodos que possuem diferença estatística são: PROD+N-Grams+Est-SWR (9.25), 2-Grams+SWR (200.25), Stem-SWR (208.375) e Stem+SWR (208.375).

A combinação PROD+N-Grams+Est-SWR usando *late fusion* obteve o melhor resultado estatístico conforme Tabela 10. Note que este resultado é relevante, pois na maioria dos trabalhos da literatura é utilizado apenas o método Stem (NEUMAYER; RAUBER, 2007; MAYER et al., 2008a, 2008b; MAYER; NEUMAYER, 2009; MAYER; RAUBER, 2010, 2011). Além disso, este resultado é convergente com os trabalhos citados, nos quais os melhores resultados obtidos foram por meio da combinação de características.

Analisando a Tabela 10 é possível verificar que o melhor método PROD+N-Grams+Est-SWR é estatisticamente similar aos demais, exceto o Stem. Desta forma, mesmo o método PROD+N-Grams+Est-SWR não atingindo o melhor resultado em todas as bases de letras de músicas, este pode ser utilizado na classificação de gêneros musicais utilizando letras de músicas. Note também que o método PROD+N-Grams+Est-SWR utiliza apenas abordagens independentes de idioma e que de forma geral pode ser utilizado em diversas bases de letras de músicas em outros idiomas.

Tabela 10: Ranking médio dos métodos utilizados

Método	Rank	Método	Rank	Método	Rank
PROD+N-Grams+Est-SWR	9.25	PROD+2-Grams+3-Grams+Stem+SWR	95.125	PROD+2-Grams+4-Grams+Stem+SWR	128.25
PROD+2-Grams+3-Grams+Est-SWR	17.0	MAX+3-Grams+4-Grams+Stem-SWR	96.0	EF+2-Grams+4-Grams+Stem+SWR	128.75
SOMA+N-Grams+Stem+Est+SWR	19.75	MAX+3-Grams+Est+SWR	96.375	MAX+3-Grams+Stem+Est-SWR	128.875
PROD+N-Grams+Est+SWR	20.875	SOMA+3-Grams+Est-SWR	97.125	MAX+3-Grams+Stem+SWR	129.125
PROD+2-Grams+3-Grams+Stem+Est+SWR	25.0	EF+3-Grams+4-Grams+Stem+SWR	98.125	SOMA+2-Grams+4-Grams+SWR	129.875
SOMA+3-Grams+4-Grams+Stem+Est+SWR	28.625	SOMA+2-Grams+Stem+Est-SWR	98.125	MAX+2-Grams+3-Grams+SWR	130.25
SOMA+3-Grams+4-Grams+Est-SWR	37.25	MAX+N-Grams+Stem-SWR	98.25	MAX+2-Grams+3-Grams+Est-SWR	131.25
PROD+2-Grams+3-Grams+Stem+Est-SWR	38.25	EF+4-Grams+Stem+SWR	98.5	MAX+2-Grams+3-Grams-SWR	133.25
PROD+4-Grams+Est-SWR	39.125	MAX+3-Grams+4-Grams+Stem+Est-SWR	99.5	PROD+2-Grams+4-Grams+Stem-SWR	133.25
PROD+3-Grams+Est+SWR	39.25	PROD+N-Grams+Stem+Est-SWR	99.625	EF+2-Grams+4-Grams-SWR	136.125
SOMA+3-Grams+4-Grams+Stem+Est-SWR	41.25	MAX+3-Grams+Est-SWR	100.125	EF+3-Grams+Stem+Est-SWR	136.25
SOMA+2-Grams+4-Grams+Stem+Est+SWR	41.75	SOMA+4-Grams+Stem+SWR	100.375	MAX+4-Grams+Stem-SWR	136.25
SOMA+N-Grams+Est+SWR	42.125	PROD+2-Grams+Est+SWR	101.875	MAX+2-Grams+4-Grams+SWR	137.875
PROD+3-Grams+Est-SWR	43.875	MAX+3-Grams+4-Grams+Stem+Est+SWR	102.25	EF+2-Grams+4-Grams+Stem-SWR	138.0
PROD+4-Grams+Est+SWR	44.75	PROD+2-Grams+3-Grams-SWR	102.5	EF+2-Grams+3-Grams+Stem+SWR	140.25
SOMA+3-Grams+Stem+Est+SWR	45.625	PROD+4-Grams+Stem+Est-SWR	104.125	EF+2-Grams+4-Grams+SWR	141.125
SOMA+3-Grams+4-Grams+Est+SWR	46.625	MAX+4-Grams+Stem+Est+SWR	105.0	SOMA+3-Grams+Stem-SWR	142.0
SOMA+2-Grams+4-Grams+Est-SWR	46.75	MAX+N-Grams+Stem+SWR	105.625	PROD+3-Grams+Stem-SWR	142.0
EF+2-Grams+4-Grams+Stem+Est+SWR	47.625	PROD+2-Grams+4-Grams-SWR	106.75	PROD+4-Grams+Stem-SWR	143.75
PROD+3-Grams+Stem+Est+SWR	47.75	SOMA+2-Grams+4-Grams+Stem-SWR	106.875	SOMA+Stem+Est+SWR	144.5
SOMA+2-Grams+4-Grams+Est+SWR	51.125	PROD+2-Grams+4-Grams+SWR	107.0	MAX+2-Grams+3-Grams+Stem-SWR	145.25
SOMA+2-Grams+4-Grams+Stem+Est-SWR	51.75	SOMA+2-Grams+4-Grams-SWR	107.875	EF+3-Grams+Est+SWR	145.625
PROD+3-Grams+4-Grams+Est+SWR	52.125	PROD+Stem+Est-SWR	109.0	4-Grams-SWR	146.0
EF+N-Grams+Stem+Est+SWR	53.0	MAX+3-Grams+Stem+Est+SWR	109.125	SOMA+Stem+Est-SWR	146.875
PROD+3-Grams+4-Grams+Stem+Est+SWR	53.125	PROD+3-Grams+Stem+SWR	109.875	4-Grams+SWR	149.125
SOMA+2-Grams+3-Grams+Est-SWR	54.25	MAX+2-Grams+3-Grams+Stem+Est+SWR	110.625	EF+2-Grams+Stem+Est+SWR	149.375
SOMA+N-Grams+Stem+Est-SWR	54.25	SOMA+2-Grams+3-Grams+Stem+SWR	110.625	EF+3-Grams+Est-SWR	150.75
SOMA+3-Grams+Stem+Est-SWR	55.0	PROD+3-Grams+4-Grams+Stem+SWR	110.75	EF+2-Grams+3-Grams+Stem-SWR	152.875
EF+3-Grams+4-Grams+Stem+Est+SWR	55.25	EF+3-Grams+4-Grams+Stem-SWR	111.375	MAX+2-Grams+Stem+Est+SWR	153.0
SOMA+4-Grams+Stem+Est+SWR	56.0	SOMA+2-Grams+Est-SWR	111.75	MAX+3-Grams+Stem-SWR	153.875
EF+4-Grams+Stem+Est-SWR	74.5	MAX+4-Grams+Stem+Est-SWR	113.125	EF+2-Grams+3-Grams-SWR	154.25
SOMA+3-Grams+4-Grams-SWR	76.125	PROD+3-Grams+4-Grams+Stem-SWR	113.875	MAX+2-Grams+Est-SWR	155.125
SOMA+3-Grams+4-Grams+Stem-SWR	76.875	PROD+2-Grams+3-Grams+Stem-SWR	114.125	EF+2-Grams+Stem+Est-SWR	158.875
SOMA+3-Grams+4-Grams+Stem+SWR	78.625	MAX+4-Grams+Stem+SWR	114.375	MAX+2-Grams+Stem+Est-SWR	159.25
EF+2-Grams+3-Grams+Stem+Est-SWR	80.0	EF+3-Grams+4-Grams-SWR	114.5	EF+2-Grams+3-Grams+SWR	161.125
MAX+3-Grams+4-Grams-SWR	80.5	MAX+2-Grams+3-Grams+Stem+SWR	115.0	MAX+2-Grams+Est+SWR	163.375
PROD+3-Grams+4-Grams+SWR	81.625	EF+3-Grams+4-Grams+SWR	116.375	PROD+2-Grams+Stem+SWR	164.875
MAX+3-Grams+4-Grams+Stem+SWR	82.125	PROD+N-Grams+Stem+SWR	116.625	EF+3-Grams+Stem+SWR	166.875
EF+2-Grams+3-Grams+Est-SWR	82.25	MAX+2-Grams+4-Grams-SWR	116.75	EF+2-Grams+Est-SWR	169.5
PROD+3-Grams+4-Grams+Est-SWR	82.625	MAX+2-Grams+4-Grams+Est-SWR	116.875	EF+2-Grams+Est+SWR	174.375
EF+2-Grams+3-Grams+Est+SWR	84.5	SOMA+4-Grams+Stem-SWR	116.875	SOMA+2-Grams+Stem+SWR	175.0
SOMA+3-Grams+4-Grams+SWR	84.5	EF+4-Grams+Stem-SWR	118.875	MAX+Stem+Est+SWR	176.375
MAX+N-Grams+Stem+Est+SWR	85.125	PROD+2-Grams+3-Grams+SWR	119.0	MAX+Stem+Est-SWR	176.5
PROD+2-Grams+4-Grams+Est+SWR	85.25	MAX+2-Grams+4-Grams+Stem+Est-SWR	119.875	3-Grams-SWR	179.5
MAX+4-Grams+Est+SWR	85.5	MAX+2-Grams+4-Grams+Stem+Est+SWR	120.375	3-Grams+SWR	179.5
MAX+N-Grams+Stem+Est-SWR	85.625	SOMA+2-Grams+Est+SWR	120.75	MAX+2-Grams+Stem+SWR	181.5
PROD+2-Grams+4-Grams+Est-SWR	85.625	PROD+Stem+Est+SWR	121.5	EF+3-Grams+Stem-SWR	182.5
MAX+3-Grams+4-Grams+Est+SWR	86.875	MAX+2-Grams+3-Grams+Est+SWR	122.5	PROD+2-Grams+Stem-SWR	187.0
EF+3-Grams+4-Grams+Est+SWR	87.5	MAX+2-Grams+4-Grams+Stem+SWR	122.75	SOMA+2-Grams+Stem-SWR	189.625
SOMA+2-Grams+4-Grams+Stem+SWR	88.25	SOMA+2-Grams+3-Grams+Stem-SWR	122.875	EF+Stem+Est+SWR	189.75
MAX+3-Grams+4-Grams+Est-SWR	88.375	SOMA+2-Grams+3-Grams-SWR	122.875	EF+2-Grams+Stem+SWR	190.5
EF+3-Grams+4-Grams+Est-SWR	90.625	SOMA+2-Grams+3-Grams+SWR	123.75	MAX+2-Grams+Stem-SWR	193.375
MAX+3-Grams+4-Grams+SWR	90.875	MAX+2-Grams+4-Grams+Est+SWR	124.875	EF+Stem+Est-SWR	194.0
EF+N-Grams+Stem+SWR	91.25	MAX+2-Grams+4-Grams+Stem-SWR	125.25	Est+SWR	198.375
EF+4-Grams+Est-SWR	91.5	MAX+2-Grams+3-Grams+Stem+Est-SWR	125.5	Est-SWR	198.75
EF+N-Grams+Stem-SWR	92.125	PROD+N-Grams+Stem-SWR	126.0	2-Grams-SWR	199.0
EF+4-Grams+Est+SWR	92.5	PROD+N-Grams+Stem+Est+SWR	126.125	EF+2-Grams+Stem-SWR	199.25
SOMA+2-Grams+Stem+Est+SWR	92.75	SOMA+3-Grams+Stem+SWR	126.25	2-Grams+SWR	200.25
MAX+N-Grams+Est+SWR	93.0	EF+3-Grams+Stem+Est+SWR	127.625	Stem-SWR	208.375
MAX+4-Grams+Est-SWR	93.625	PROD+4-Grams+Stem+SWR	127.625	Stem+SWR	208.375

7 CONCLUSÕES

A classificação de gêneros musicais utilizando letras de músicas em diferentes idiomas ainda é um campo de estudo pouco investigado e que carece de ferramentas de PLN devido ao enfoque dado a ferramentas dependentes de idioma (normalmente o Inglês). Além disso, a utilização de letras de músicas pode auxiliar e em alguns casos substituir outros recursos utilizados na classificação de gêneros musicais. Desta forma, neste trabalho é realizada uma investigação de qual abordagem é mais adequada para a tarefa de classificação automática de gêneros musicais ao lidar com letras de músicas em diferentes idiomas.

Neste trabalho foram empregadas técnicas como *stemming* e n-gramas. A técnica *stemming* usa recursos dependentes de idioma que podem não estar disponíveis para alguns idiomas enquanto a abordagem n-gramas provê uma abordagem independente de idioma. Em adição a estas abordagens, também foi realizado um estudo do impacto das características estilísticas de texto.

Para avaliar as abordagens dependentes e independentes de idioma foram criadas três novas bases de dados de letras de músicas: *Nordic Lyrics Database* (NLD), *Latin Music Lyrics Database* (LMLD) e *Lyrics GTZAN* (LGTZAN). Estas três bases de dados foram combinadas para criar a *Joint Lyrics Database* (JLD).

Os melhores resultados *no fusion* para cada base de dados foram 87.5% para a NLD com 4-gramas com SWR; 65.6% para a LMLD com a abordagem 4-gramas sem SWR; 43.9% para a LGTZAN com 4-gramas sem SWR e 52.2% para a JLD com 3-gramas com SWR. Note que a abordagem 4-gramas sem SWR obteve 65.6% para a LMLD e 43.9% para a LGTZAN. Portanto, com base na análise dos experimentos computacionais realizados neste trabalho, a recomendação é usar 4-gramas ao lidar com letras em diferentes idiomas. A abordagem n-grama também tem a vantagem de ser independente de idioma.

Adicionalmente foram realizados experimentos para avaliar o impacto da remoção das *stopwords* nas técnicas de extração de características avaliadas individualmente na tarefa

de classificação de gêneros musicais utilizando letras de músicas. Isto é importante por duas razões: Primeiro, não está claro se as *stopwords* deveriam ser removidas ao lidar com letras de músicas; Segundo, as *stopwords* são recursos dependentes de idiomas e podem não estar disponíveis para alguns idiomas. Entretanto, o uso ou não das *stopwords* depende da representação a ser utilizada. Para as características estilísticas, é evidente que o uso de SWR sempre melhora os resultados usando a abordagem *early fusion*. Para outras abordagens os resultados não foram claros. Por exemplo, o uso de SWR para os experimentos na NLD dificulta a performance de *stemming* e n-gramas, enquanto melhora a performance da representação n-gramas na JLD.

Outra contribuição deste trabalho foi verificar se é benéfico ou não combinar diferentes características. A análise dos resultados experimentais sugerem que estas devem ser utilizadas com outras representações como características estilísticas, pois usualmente melhoram os resultados da classificação.

A última contribuição deste trabalho foi verificar a abordagem *late fusion* com os seguintes métodos: probabilidade máxima, soma da probabilidade e produto da probabilidade. Analisando todos os resultados das quatro bases de dados de letras de músicas utilizadas neste trabalho verificou-se que as melhores abordagens globais foram N-Gramas+*Stemming*+SWR e N-Gramas+*Stemming*+Estilística+SWR usando *early fusion* na NLD, em que ambas atingiram 95.5% de *F-measure*. Além disso, nas bases de dados LGTZAN e JLD, por exemplo, o melhor resultado foi atingido com a combinação 2-Gramas+3-Gramas+Estilística-SWR (78.5% para a LGTZAN e 71.5% para a JLD) usando *late fusion* com o produto das probabilidades, sendo esta uma representação com alta dimensionalidade. Neste sentido, visando verificar estatisticamente estes resultados obtidos, os métodos foram avaliados utilizando os testes de Friedman e Nemenyi. Nestes testes estatísticos foi verificado que o melhor método foi PROD+N-Gramas+Est-SWR, o qual é estatisticamente similar aos melhores resultados obtidos em todas as quatro bases. Portanto, o método PROD+N-Gramas+Est-SWR é o mais indicado para a classificação gêneros musicais utilizando letras de músicas em diferentes idiomas, pois todas as representações combinadas são independentes de idiomas e podem ser aplicadas em outras bases de letras de músicas. Além disso, é importante ressaltar que o método PROD+N-Gramas+Est-SWR não foi utilizado anteriormente em outros trabalhos da literatura, em que apenas o *stemming* foi aplicado, o qual é um método dependente de idioma.

Neste trabalho foram feitas várias combinações das técnicas de mineração de texto aplicadas a classificação de gêneros musicais utilizando letras de músicas. Porém, não foram realizadas comparações com outros classificadores, tendo em vista que o foco deste trabalho

era a verificação da combinação das técnicas por meio de *early fusion* e *late fusion*, as quais não foram realizadas na literatura, para o melhor do conhecimento dos autores. No entanto, esta verificação da combinação das técnicas de classificação de gêneros musicais utilizando letras de músicas com diferentes classificadores pode ser uma análise interessante e melhorar, inclusive, os resultados reportados neste trabalho. Como direções futuras, serão utilizados outros classificadores e serão criadas outras bases de dados de letras de músicas em outros idiomas. Além disso, pretende-se utilizar informação hierárquica nas bases de dados e testar as combinações feitas neste trabalho em um contexto de classificação hierárquica de gêneros musicais utilizando letras de músicas.

REFERÊNCIAS

- BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. **Data mining and knowledge discovery**, Springer, v. 2, n. 2, p. 121–167, 1998.
- CAVNAR, W. B.; TRENKLE, J. M. N-gram-based text categorization. In: **Proc. of the 3rd Annual Symposium on Document Analysis and Information Retrieval**. 1994. p. 161–175.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **The Journal of Machine Learning Research**, Journal of Machine Learning Research, v. 7, p. 1–30, 2006.
- DHANARAJ, R.; LOGAN, B. Automatic prediction of hit songs. In: **Proc. of the 6th Int. Conf. on Music Information Retrieval (ISMIR)**. 2005. p. 488–491.
- DOWNIE, J. S. The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research. **Acoustical Science and Technology**, v. 29, n. 4, p. 247–255, 2008.
- FABBRI, F. **Browsing Music Spaces: Categories And The Musical Mind**. 1999. Disponível em: <<http://www.mediamusicstudies.net/tagg/xpdfs/ffabbri9907us.pdf>>.
- FABBRI, F. Browsing music spaces: Categories and the musical mind. In: **Proc. of the IASPM Conference**. 1999.
- FELL, M.; SPORLEDER, C. Lyrics-based analysis and classification of music. In: **Proc. of the 25th Int. Conf. on Computational Linguistics: Technical Papers**. 2014. p. 620–631.
- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. **Journal of the american statistical association**, Taylor & Francis, v. 32, n. 200, p. 675–701, 1937.
- FRIEDMAN, M. A comparison of alternative tests of significance for the problem of m rankings. **The Annals of Mathematical Statistics**, JSTOR, v. 11, n. 1, p. 86–92, 1940.
- FU, Z.; LU, G.; TING, K. M.; ZHANG, D. A survey of audio-based music classification and annotation. **IEEE Transactions on Multimedia**, IEEE, v. 13, n. 2, p. 303–319, 2011.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. **ACM SIGKDD explorations newsletter**, ACM, v. 11, n. 1, p. 10–18, 2009.
- HIRJEE, H.; BROWN, D. Using automated rhyme detection to characterize rhyming style in rap music. *Empirical Musicology Review*, 2010.
- HO, T. K.; HULL, J. J.; SRIHARI, S. N. Decision combination in multiple classifier systems. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 16, n. 1, p. 66–75, 1994.

- HU, X.; DOWNIE, J. S.; EHMAN, A. F. Lyric text mining in music mood classification. In: **Proc. of the 10th Int. Conf. on Music Information Retrieval (ISMIR)**. 2009. p. 411–416.
- HU, X.; DOWNIE, S. J. Improving mood classification in music digital libraries by combining lyrics and audio. In: **Proc. of the 10th Annual Joint Conf. on Digital Libraries**. 2010. p. 159–168.
- HU, X.; DOWNIE, S. J. When lyrics outperform audio for music mood classification: a feature analysis. In: **Proc. of the 11th Int. Conf. on Music Information Retrieval (ISMIR)**. 2010. p. 619–624.
- HU, Y.; CHEN, X.; YANG, D. Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. In: **Proc. of the 10th Int. Conf. on Music Information Retrieval (ISMIR)**. 2009. p. 123–128.
- KITTLER, J.; HATEF, M.; DUIN, R. P.; MATAS, J. On combining classifiers. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 20, n. 3, p. 226–239, 1998.
- KOHAVI, R.; PROVOST, F. Glossary of terms. **Machine Learning**, v. 30, n. 2-3, p. 271–274, 1998.
- LAROSE, D. T. **Discovering knowledge in data: an introduction to data mining**. : John Wiley & Sons, 2014.
- LAURIER, C.; GRIVOLLA, J.; HERRERA, P. Multimodal music mood classification using audio and lyrics. In: **Proc. of the 7th Int. Conf. on Machine Learning and Applications**. 2008. p. 688–693.
- LEE, J. H.; DOWNIE, J. S. Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In: **Proc. of the 5th Int. Conf. on Music Information Retrieval (ISMIR)**. 2004. p. 441–446.
- LI, T.; OGIHARA, M.; LI, Q. A comparative study on content-based music genre classification. In: **ACM. Proc. of the 26th annual international ACM SIGIR conference on Research and development in information retrieval**. 2003. p. 282–289.
- LIMA, A. A. de; NUNES, R. M.; RIBEIRO, R. P.; SILLA JR., C. N. Nordic music genre classification using song lyrics. In: **Proc. of the 19th Int. Conf. on Natural Language Processing and Information Systems**. 2014. (Lecture Notes in Computer Science, v. 8455), p. 89–100.
- MARTINDALE, C. **Romantic progression: The psychology of literary history**. : Hemisphere publishing corporation, 1975.
- MARTINDALE, C. **The clockwork muse: The predictability of artistic change**. : Basic Books, 1990.
- MAYER, R.; NEUMAYER, R. Multi-modal analysis of music: A large-scale evaluation. In: **Proc. of the Workshop on Exploring Musical Information Spaces**. 2009. p. 30–35.

MAYER, R.; NEUMAYER, R.; RAUBER, A. Combination of audio and lyrics features for genre classification in digital audio collections. In: **Proc. of the 16th ACM Int. Conf. on Multimedia**. 2008. p. 159–168.

MAYER, R.; NEUMAYER, R.; RAUBER, A. Rhyme and style features for musical genre classification by song lyrics. In: **Proc. of the 9th Int. Conf. on Music Information Retrieval (ISMIR)**. 2008. p. 337–342.

MAYER, R.; RAUBER, A. Building ensembles of audio and lyrics features to improve musical genre classification. In: **Proc. of the Int. Conf. on Distributed Framework and Applications**. 2010. p. 1–6.

MAYER, R.; RAUBER, A. Musical genre classification by ensembles of audio and lyrics features. In: **Proc. of the 12th Int. Conf. on Music Information Retrieval (ISMIR)**. 2011. p. 675–680.

MCKAY, C.; BURGOYNE, J. A.; HOCKMAN, J.; SMITH, J. B. L.; VIGLIENSONI, G.; FUJINAGA, I. Evaluating the genre classification performance of lyrical features relative to audio, symbolic and cultural features. In: **Proc. of the 11th Int. Conf. on Music Information Retrieval (ISMIR)**. 2010. p. 213–218.

MCKAY, C.; FUJINAGA, I. Musical genre classification: Is it worth pursuing and how can it be improved? In: **Proc. of the 7th Int. Conf. on Music Information Retrieval (ISMIR)**. 2006. p. 101–106.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas Inteligentes-Fundamentos e Aplicações**, v. 1, p. 1, 2003.

NEMENYI, P. Distribution-free multiple comparisons. In: INTERNATIONAL BIOMETRIC SOCIETY. **Biometrics**. 1962. v. 18, n. 2, p. 263.

NEUMAYER, R.; RAUBER, A. Integration of text and audio features for genre classification in music information retrieval. In: **Advances in Information Retrieval**. : Springer, 2007. p. 724–727.

ORIO, N. Music retrieval: a tutorial and review. **Foundations and Trends in Information Retrieval**, now publishers Inc, v. 1, n. 1, p. 1–90, 2006.

PACHET, F.; AUCOUTURIER, J. Improving timbre similarity: How high is the sky? **Journal of negative results in speech and audio sciences**, v. 1, n. 1, p. 1–13, 2004.

PLATT, J. C. Fast training of support vector machines using sequential minimal optimization. **Advances in kernel methods**, p. 185–208, 1999.

PORTER, M. F. An algorithm for suffix stripping. **Program: electronic library and information systems**, MCB UP Ltd, v. 14, n. 3, p. 130–137, 1980.

RIBEIRO, R. P.; ALMEIDA, M. A. P.; SILLA JR., C. N. The ethnic lyrics fetcher tool. **EU-RASIP Journal on Audio, Speech, and Music Processing**, Springer, v. 2014, n. 1, p. 1–10, 2014.

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. **Information processing & management**, Elsevier, v. 24, n. 5, p. 513–523, 1988.

SCARINGELLA, N.; ZOIA, G.; MLYNEK, D. Automatic genre classification of music content: a survey. **IEEE Signal Processing Magazine**, IEEE, v. 23, n. 2, p. 133–141, 2006.

SEBASTIANI, F. Machine learning in automated text categorization. **ACM computing surveys**, v. 34, n. 1, p. 1–47, 2002.

SHUYO, N. **Language Detection Library for Java**. 2010. Disponível em: <<http://code.google.com/p/language-detection/>>.

SILLA JR., C. N.; FREITAS, A. A. Novel top-down approaches for hierarchical classification and their application to automatic music genre classification. In: **IEEE Int. Conf. on Systems, Man and Cybernetics (SMC)**. 2009. p. 3499–3504.

SILLA JR., C. N.; FREITAS, A. A. A survey of hierarchical classification across different application domains. **Data Mining and Knowledge Discovery**, Springer, v. 22, n. 1-2, p. 31–72, 2011.

SILLA JR., C. N.; KOERICH, A. L.; KAESTNER, C. A. A. The latin music database. In: **Proc. of the 9th Int. Conf. on Music Information Retrieval (ISMIR)**. 2008. p. 451–456.

SILLA JR., C. N.; KOERICH, A. L.; KAESTNER, C. A. A. Improving automatic music genre classification with hybrid content-based feature vectors. In: **Proc. of the ACM Symp. on Applied Computing**. 2010. p. 1702–1707.

SNOEK, C. G.; WORRING, M.; SMEULDERS, A. W. Early versus late fusion in semantic video analysis. In: **ACM. Proc. of the 13th annual ACM international conference on Multimedia**. 2005. p. 399–402.

TOKUNAGA, T.; MAKOTO, I. **Text categorization based on weighted inverse document frequency**. 1994.

WHITMAN, B.; SMARAGDIS, P. Combining musical and cultural features for intelligent style detection. In: **Proc. of the 3rd Int. Conf. on Music Information Retrieval (ISMIR)**. 2002. p. 47–52.

WU, H.; SALTON, G. A comparison of search term weighting: Term relevance vs. inverse document frequency. In: **Proc. of the 4th Special Interest Group on Information Retrieval**. 1981. p. 30–39. ISSN 0163-5840.

XU, L.; KRZYŻAK, A.; SUEN, C. Y. Methods of combining multiple classifiers and their applications to handwriting recognition. **IEEE Transactions on Systems, man and cybernetics**, IEEE, v. 22, n. 3, p. 418–435, 1992.

ZAANEN, M. v.; KANTERS, P. Automatic mood classification using tf*idf based on lyrics. In: **Proc. of the 11th Int. Conf. on Music Information Retrieval (ISMIR)**. 2010. p. 75–80.