Nordic Music Genre Classification Using Song Lyrics

Adriano A. de Lima, Rodrigo M. Nunes,
Rafael P. Ribeiro, and Carlos N. Silla Jr.

Computer Music Technology Laboratory
Federal University of Technology of Parana (UTFPR)
Av. Alberto Carazzai, 1.640 - 86300-000
Cornélio Procópio, PR, Brazil
{adrianolima.aal,kyllopardiun,rpr.rafa,carlos.sillajr}@gmail.com
http://music.cp.utfpr.edu.br

Abstract. Lyrics-based music genre classification is still understudied within the music information retrieval community. The existing approaches, reported in the literature, only deals with lyrics in the English language. Thus, it is necessary to evaluate if the standard text classification techniques are suitable for lyrics in languages other than English. More precisely, in this work we are interested in analyzing which approach gives better results: a language-dependent approach using stemming and stopwords removal or a language-independent approach using n-grams. To perform the experiments we have created the Nordic music genre lyrics database. The analysis of the experimental results shows that using a language-independent approach with the n-gram representation is better than using a language-dependent approach with stemming. Additional experiments using stylistic features were also performed. The analysis of these additional experiments has shown that using stylistic features combined with the other approaches improve the classification results.

Keywords: Lyrics Classification, Multi-language text classification, Music Genre Classification.

1 Introduction

Multimedia technologies have been an important field in many research areas, for example, Music Information Retrieval (MIR) has focused in important research directions such as music similarity retrieval, musical genre classification or music analysis and knowledge representation [1]. The content of the multimedia data might be audio, album covers, social and cultural data, music videos or song lyrics. In [2–6] it has been shown that song lyrics are relevant to assist or substitute other approaches for the task of music genre classification using standard text classification techniques. However, according to [7] the use of lyrics generally depends on language processing tools, that might not be available for songs whose lyrics are not in English (e.g. for Nordic music).

E. Métais, M. Roche, and M. Teisseire (Eds.): NLDB 2014, LNCS 8455, pp. 89–100, 2014.

The main objective of this paper is to verify which approach is more suitable for the lyrics-based music genre classification when dealing with lyrics that are not in the English language. For this purpose we have employed two standard approaches for lyrics processing (presented in Section 2): the first approach uses language-dependent resources while the second approach is language-independent. In order to evaluate both approaches on non-English Lyrics, we developed the Nordic Music Genre Lyrics Database (presented in Section 3). Our computation experiments with the two different approaches for the task of lyrics-based music genre classification and their results are shown in Section 4. In Section 5 we present the conclusions and future research directions of this work.

2 Lyrics Processing

An approach commonly employed for text processing is bag-of-words wherein each document is represented as word vectors that occurs into document, this approach is necessary to standardize the documents into a fixed format understood by the classifier. Thus, in this section we present the text processing techniques used to build the bag-of-words model applied in this work for the task of lyrics-based music genre classification. In Section 2.1 we describe the lyrics pre-processing steps. In Section 2.2 we describe the optional step of stopwords removal. Note that although stopwords are language-dependent, we have allowed this optional step in both approaches for the sake of completeness in our experiments. In Section 2.3 we describe the n-gram approach, which is a language-independent approach, for representing documents as word vectors. In Section 2.4 we describe the language-dependent stemming approach for representing documents as word vectors. In Section 2.5 we provide a brief explanation of the Term Frequency-Inverse Document Frequency (TF-IDF) technique and how it is used with the approaches presented in sections 2.3 and 2.4. In Section 2.6 we present an alternative approach for extracting features from lyrics, known as stylistic features.

2.1 Pre-processing

This step removes accents, punctuation, special characters, breaking words with hyphens, treatment of capital letters and equivalence of characters process. This procedure is necessary because we are dealing with non-usual lyrics, special attention is needed to some special characters present in the Nordic lyrics that should be mapped to specific characters in order to maintain their meaning. Table 1 presents the mapping between the Nordic languages special characters and their equivalent form.

2.2 Stopwords Removal

The stopwords are very common words (i.e. articles and prepositions) that carry little semantic meaning to the text, these irrelevant words are included in a

Table 1. Special Characters Mapping for the Nordic Languages

Language(s)	Special Character	Equivalent Character
Danish, Norwegian	æ	ae
Danish, Norwegian	ø	oe
Danish, Norwegian, Swedish	å	aa
Swedish	ö	oe
Swedish	ä	ae

pre-defined list, called stopwords list or stoplist[1] [8]. In this work we have used one stoplist for each Nordic language with 114 stopwords for Swedish, 176 for Norwegian and 94 for Danish.

In [9, 10] the authors studied the impact of stopwords for text and music mood classification. However, as we are interested in understanding the impact of these stopwords in the music genre classification using Nordic lyrics, we have analyzed whether it is advantageous or not to remove them before performing the task of lyrics-based music genre classification.

2.3 N-Grams

This technique transforms words into n-grams. A n-gram is a sequence of n characters (adding the special character "_" to denote the beginning and ending of the word), for example, given the term "text": for n=2, the representative attributes would be "_t", "te", "ex", "xt" and "t_"; for n=3, "_te", "tex", "ext" and "xt_"; and for n=4, "_tex", "text" and "ext_" [11].

In distinct language cases this technique is generally used when languages have a common root. Thus, the variation of the some words does not influence the morphemes that are common between languages. Therefore, this approach allows to infer the morphemes dividing the word into a character sequence of fixed size, n, which is predefined according with application context.

2.4 Stemming

The stemming approach is used to obtain the root form of all the words from a set of lyrics. For example, given the words "computing" and "computer", the result of a stemming algorithm could be "comput" [12]. In [13] it has been shown that this technique is largely used in the text classification domain with the objective of reducing the dimensionality. It is also a very fast procedure. It is important to highlight that the stemming algorithms are language specific and therefore may not be available to some languages [14].

2.5 TF-IDF

There are many methods to calculate the weight of each term (e.g., Weighted Inverse Document Frequency (WIDF) [15] and Term Relevance [16]), being the

[1] The stoplists were obtained from `http://snowball.tartarus.org/`

TF-IDF one of most commonly employed [17]. The advantage of TF-IDF, is the trade-off between how often a term occurs (Term frequency) and in how many documents the term occurs. The rationale being that a term that occurs frequently in a given document may be important but only if it does not occur in every other document as well. The formula for computing the TF-IDF is presented in Equation (1).

$$TFIDF = TF \times \log \frac{nD}{nDF} \tag{1}$$

where:

- TF is frequency of the term in a document, in other words, the number times that term occurs in a document;
- nD is the number of documents used;
- nDF is the number of documents where occurs the representation of that term;

2.6 Stylistic Features

The use of stylistic features for lyrics-based music genre classification was originally proposed in the work of [2]. In this work they have proposed simple counts of special punctuations (e.g., "!", "-", "?"), occurrences of digits, words and unique words per line, unique words ratio and characters per word. The main motivation behind using stylistic features is to capture the "interjection words" in the lyrics,

Table 2. Subset of the stylistic features proposed in [10] used in this work

Feature	Definition
interjection words	normalized frequency of "hey", "ooh", "yo", "uh", "ah", "yeah", "hej", "vad", "hejsan", "oj", "aj", "ja", "hei"
special punctuation marks	normalized frequency of "!", ":", "?", "-", ";"
NUMBER	normalized frequency of all non-year numbers
No. of words	total number of words
No. of uniqWords	total number of unique words
repeatWordRatio	(No. of words - No. of uniqWords) / No. of words
avgWordLength	average number of characters per word
No. of lines	total number of lines
No. of uniqLines	total number of unique lines
No. of blankLines	number of blank lines
blankLineRatio	No. of blankLines / No. of lines
avgLineLength	No. of words / No. of lines
stdLineLength	standard deviation of number of words per line
uniqWordsPerLine	No. of uniqWords / No. of lines
repeatLineRatio	(No. of lines - No. of uniqLines) / No. of lines
avgRepeatWordRatioPerLine	average repeat word ratio per line
stdRepeatWordRatioPerLine	standard deviation of repeat word ratio per line

special punctuations and to compute simple text statistics. In this work we employ a subset of the stylistic features proposed for the task of lyrics-based music mood classification in [10]. We have used this subset because it complements the original stylistic features with new measures and also due to the improved results obtained in [2]. Note that in this work we have not used the stylistic features related to the song recordings (e.g. the number of words per minute or the number of lines per minute). Table 2 presents the stylistic features used in this work.

The main advantage of the stylistic features is that they use simple statistical measures based on word or character frequencies. In the context of this work, this technique was used in order to evaluate its suitability for the task of lyrics-based music genre classification.

3 Experimental Settings

In this section we present the process to create lyrics database with its characteristics (Sections 3.1 and 3.2) and the classifier used in the experiments (Section 3.3).

3.1 Nordic Lyrics Database Creation

The Nordic Music Genre Classification Lyrics Database was created using the following procedure: In the first stage we applied a questionnaire to some residents of Denmark, Norway and Sweden. In this questionnaire we asked the participants to inform their favorite songs along with the songs title, performing artist, music genre and language. Note that we imposed a restriction on the questionnaire that only songs sung in either Danish, Swedish or Norwegian were accepted. In the second stage, we searched the web manually to acquire the lyrics for our experiments. The main motivation behind the development of this novel database is to extend the analysis of lyric-based music genre classification for languages other than English.

The final version of the Nordic Music Genre Classification Lyrics Database contains 1,513 song lyrics divided in five genres (Dance: 265, Pop Rock: 263, Rap: 100, Pop: 357 and Rock: 528) distributed in three Nordic languages (Fig. 1 shows the distribution of song lyrics according to each language while Figures 2 to 6 show the word clouds for each genre using the 150 words with the highest values of TF-IDF and a list of the top 10 terms translated into the English language.). The music genres were labeled following the definition shown in [18]: “a kind of music, as it is acknowledged by a community for any reason or purpose or criteria”. It is noteworthy that in the Nordic lyrics they contain some terms from the English language.

3.2 Dataset Details

The quantity of attributes generated by representation for Nordic Music Genre Classification Lyrics Database using stopwords (represented by Sw) or after of stopwords removal are shown in Table 3.

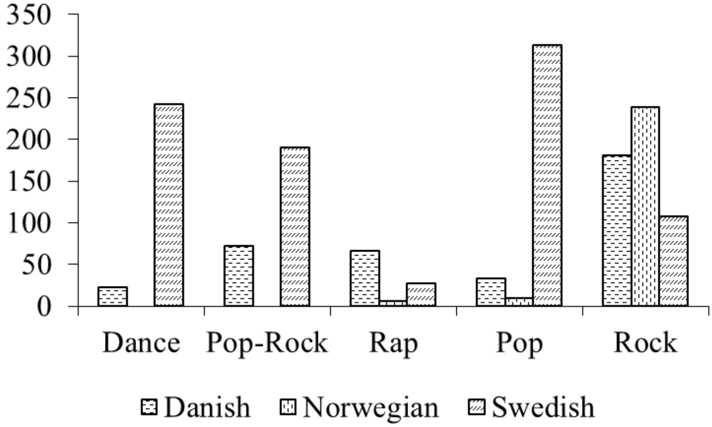

Fig. 1. Distribution of song lyrics by genres according to each Nordic language

Term	TF-IDF	Translation
jag	0.153	i (swedish)
aer	0.150	to be (verb)
i	0.137	in/of/at (preposition)
du	0.124	you (personal pronoun)
foer	0.121	before
och	0.120	and (swedish)
ska	0.120	shall
igen	0.107	again
vill	0.107	will
naer	0.105	near or when/that/as

Fig. 2. Word cloud with the 150 words with the highest TF-IDF and a list of the top 10 terms translated into the English language for the Pop music genre

Term	TF-IDF	Translation
aer	0.196	to be (verb)
i	0.155	in/of/at (preposition)
jag	0.147	i (personal pronoun)
du	0.145	you (personal pronoun)
vi	0.129	we (personal pronoun)
och	0.128	and
dom	0.118	those/them (swedish)
foer	0.111	before
en	0.111	a/an (article)
saa	0.111	so/then (adverb) or such (adjective)

Fig. 3. Word cloud with the 150 words with the highest TF-IDF and a list of the top 10 terms translated into the English language for the Pop Rock music genre

Term	TF-IDF	Translation
aer	0.181	to be (verb)
jag	0.151	i (swedish)
saa	0.149	so/then or such
i	0.148	in/of/at (preposition)
eg	0.140	i (new norwegian)
e	0.138	to be (dialect form)
va	0.130	to be (informal)
det	0.117	the (article) or it/that (prenom)
jeg	0.117	i (norwedian/danish)
te	0.107	to (norwedian dialect)

Fig. 4. Word cloud with the 150 words with the highest TF-IDF and a list of the top 10 terms translated into the English language for the Rock music genre

Term	TF-IDF	Translation
i	0.181	in/of/at
jeg	0.163	i (personal pronoun)
bellstar	0.163	belle star (soap opera star)
baby	0.130	baby ("pretty" girl)
ah	0.126	onomatopoeia
saa	0.126	so/then or such
aer	0.126	to be (verb)
paa	0.117	on/at/in/upon/ into (preposition)
og	0.109	and
naar	0.097	when

Fig. 5. Word cloud with the 150 words with the highest TF-IDF and a list of the top 10 terms translated into the English language for the Rap music genre

Term	TF-IDF	Translation
aer	0.162	to be (verb)
jag	0.141	i (swedish)
du	0.135	you (personal pronoun)
i	0.133	in/of/at (preposition)
och	0.125	and (swedish)
ich	0.118	i (german)
vill	0.115	will
foer	0.107	before
dej	0.106	you (as an object in the setence)
blue	0.102	blue (english)

Fig. 6. Word cloud with the 150 words with the highest TF-IDF and a list of the top 10 terms translated into the English language for the Dance music genre

Table 3. Number of attributes by representation

Representation	Number of Attributes	
	With Sw	Without Sw
Stemming	20,962	20,704
2-Grams	1,087	1,079
3-Grams	9,945	9,829
4-Grams	36,741	35,705

An analysis of the different representations presented in Table 3 shows that the number of attributes grows considerably as we increase the parameter n for the n-grams representation. The stemming representation contains more than 20,000 features which is a high dimensionality when compared to the bigram and trigram representations. Furthermore, according to the details presented in Table 3, the optional step of stopwords removal has little impact in filtering out the common words in Nordic Lyrics. This might be due to the fact that music composers aims at creating new song lyrics by using seldom used words. The stylistic features on other hand produce a vector of low dimensionality with 17 only numeric values that can be used alone or in conjunction with other attributes.

3.3 Classifier

In this work we have employed the Support Vector Machine (SVM) classifier for the task of lyrics-based music genre classification. The SVM classifier is a supervised machine learning algorithms that uses hyperplanes to separate two classes by mapping the training examples [19]. With the objective of classify multiple classes we use the Sequential Minimal Optimization (SMO) algorithm[2] for training SVMs with default parameters and normalized features [21].

In [22] it has been shown that SVMs present three main advantages: they do not require complex tuning of parameters, they can generalize small training set and they are suited to learning in high dimensional spaces. Furthermore, the SVM classifier is gaining attention in MIR mainly for its performance and better results compared with other standard classifiers such as Decision Trees (DT), K-Nearest Neighbors (KNN) and Random Forests (RF) [23]. Among other papers that used SVM and its derivations for music genre classification task are [2–6].

4 Computational Results

In this section we are interested in answering the following questions by using controlled experiments: When dealing with Nordic lyrics which document representation should be used (Section 4.1)? Is a language-dependent representation

[2] Implementation from the Weka Data Mining Tool [20].

(using language tools such as stemmers) better than language-independent representation (using n-grams)? What is the impact of stopwords in music genre classification (Section 4.2)? Is it possible to combine the stylistic features with the other representations to obtain better results (Section 4.3)? For these purposes we perform our experiments using the Nordic lyrics music genre classification dataset presented in Section 3.1. To evaluate the results in this work we have used the F_1-score (which is the harmonic mean of precision and recall measures) with the ten-fold cross-validation procedure.

4.1 Which Representation to Use?

The results for the individual representations without removing the stopwords are presented in Table 4. Analyzing the results we had three main findings: First, the n-gram representation is better than stem in all parametrization. This is an interesting result as the n-gram representation can be used with any language without needing any language-specific resources. Second, by increasing the value of n in the n-gram representation it is possible to improve the classification results. Third, the use of stylistic features provides an interesting result as it achieves similar classification results with the stemming approach but using only 17 features.

Table 4. Results for the individual representations without stopwords removal

Music Genre	Stem	2-Grams	3-Grams	4-Grams	Stylistic
Dance	13.4	42.0	66.1	75.7	43.4
Pop Rock	14.8	62.5	67.1	100	00.7
Rap	47.2	44.0	97.5	47.4	48.6
Pop	37.4	63.0	77.0	80.1	21.1
Rock	62.5	74.7	84.5	98.8	56.3
Overall	38.7	62.9	77.4	87.1	35.6

4.2 What Is the Impact of Stopwords in Lyrics Classification?

When dealing with lyrics is still not clear if the stopwords should be removed. For this reason, we present the results for the individual representations with stopwords removal in Table 5. The analysis of the results presented in Table 5 shows that removing the stopwords from the lyrics harms the classification results for almost every representation.

4.3 Combining the Stylistic Features with the Other Representations

In this subsection we are interested in verifying if it is possible to improve the classification results by combining the stylistic features with the other representations. The main motivation for this combination is that the stylistic features

Table 5. Results for the individual representations with stopwords removal

Music Genre	Stem	2-Grams	3-Grams	4-Grams	Stylistic
Dance	10.5	42.1	68.4	75.3	43.3
Pop Rock	14.9	59.8	67.5	100	00.7
Rap	40.3	39.8	97.5	50.6	50.5
Pop	31.4	64.7	78.3	80.7	09.9
Rock	59.2	73.8	84.7	99.0	55.0
Overall	35.1	62.6	78.4	87.5	32.6

obtained individual results that were on par with the results obtained by using the stemming representation. Furthermore, the stylistic features are a language independent approach that can be easily computed and has only 17 attributes. The results of this experiment is presented in Table 6.

The analysis of the results in Table 6 shows that by combining the stylistic features with the stemming representation the classification results was improved from 38.7% (using only stemming) to 47%. This result is still worse than using any of the n-grams representation alone. Furthermore, when combining the n-grams representation with the stylistic features there is a small gain in classification results. More precisely from 62.9% to 64.2% for bigrams; from 77.4% to 78.1% for trigrams and from 87.1% to 87.9% for quadgrams.

Table 6. Results for the combination of the stylistic features with the other representations and without stopwords removal

Music Genre	STY+Stem	STY+2-Grams	STY+3-Grams	STY+4-Grams
Dance	38.6	51.1	68.8	74.5
Pop Rock	22.2	60.3	66.9	100
Rap	50.9	47.3	97.5	57.0
Pop	43.6	65.2	78.3	81.3
Rock	65.0	75.4	84.6	99.1
Overall	47.0	64.2	78.1	87.9

5 Conclusions

In this paper we have investigated which text classification approach is more suitable for the task of non-English lyrics-based music genre classification. In our experiments we have employed the stemming and n-grams based approaches. The stemming approach uses language-dependent resources which may not be available for some languages while using n-grams provides a language-independent approach. In addition to these approaches we have also studied the impact of using stylistic features.

In order to evaluate these approaches on non-English lyrics we developed a novel dataset. This dataset is the Nordic lyrics music genre classification dataset and it contains 1,513 lyrics from three different Nordic languages (Swedish, Norwegian and Danish) and from five different music genres (Dance, Pop Rock, Rap, Pop and Rock).

Our experimental results with the Nordic lyrics music genre classification dataset suggests that the use of the n-grams approach is more suitable for the task than the use of the stemming approach or the stylistic features alone. This is an interesting result as the n-grams approach is a language independent approach. Furthermore, we have performed experiments with bigrams, trigrams and quadgrams and the best results were obtained by using the quadgrams representation.

Additional experiments were also performed in order to evaluate the impact of stopwords removal in the task of lyrics classification. This is important for two reasons: First, it is not clear if stopwords should be removed when dealing with lyrics; Second, stopwords are a language specific resource and they may not be available for some languages. Based on our experimental results our suggestion is that the step of stopword removal is not performed when dealing with lyrics-based music genre classification.

The last contribution of the paper was to verify whether it was beneficial or not to combine the stylistics features with the other approaches used in this work. The analysis of our experimental results suggests that they should be used with other representations as the stylistic features always improve the classification results.

As a future research directions we plan on creating other non-English lyrics datasets in order to perform extensive experiments in the task of non-English lyrics-based music genre classification and also to perform additional experiments using other text classification features, such as the ones used in [24, 25].

References

1. Orio, N.: Music retrieval: a tutorial and review. Foundations and Trends in Information Retrieval 1(1), 1–90 (2006)
2. Mayer, R., Neumayer, R., Rauber, A.: Combination of audio and lyrics features for genre classification in digital audio collections. In: Proceedings of the 16th ACM International Conference on Multimedia, pp. 159–168 (2008)
3. Mayer, R., Neumayer, R., Rauber, A.: Rhyme and style features for musical genre classification by song lyrics. In: Proceedings of the 9th International Conference on Music Information Retrieval, pp. 337–342 (2008)
4. Mayer, R., Neumayer, R.: Multi-modal analysis of music: A large-scale evaluation. In: Proceedings of the Workshop on Exploring Musical Information Spaces, pp. 30–35 (2009)
5. Mayer, R., Rauber, A.: Building ensembles of audio and lyrics features to improve musical genre classification. In: Proceedings of the International Conference on Distributed Framework and Applications, pp. 1–6 (2010)
6. Mayer, R., Rauber, A.: Musical genre classification by ensembles of audio and lyrics features. In: Proceedings of International Conference on Music Information Retrieval, pp. 675–680 (2011)

7. Silla Jr., C.N., Koerich, A.L., Kaestner, C.A.A.: Improving automatic music genre classification with hybrid content-based feature vectors. In: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 1702–1707 (2010)
8. El-Khair, I.A.: Effects of stop words elimination for arabic information retrieval: a comparative study. International Journal of Computing & Information Sciences 4(3), 119–133 (2006)
9. Yu, B.: An evaluation of text classification methods for literary study. Literary and Linguistic Computing 23(3), 327–343 (2008)
10. Hu, X., Downie, J.S.: Improving mood classification in music digital libraries by combining lyrics and audio. In: Proceedings of the 10th Annual Joint Conference on Digital Libraries, pp. 159–168 (2010)
11. Cavnar, W.B., Trenkle, J.M.: N-gram-based text categorization. In: Proceedings of the 3rd Annual Symposium on Document Analysis and Information Retrieval, pp. 161–175 (1994)
12. Porter, M.F.: An algorithm for suffix stripping. Program: Electronic Library and Information Systems 14, 130–137 (1980)
13. Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1–47 (2002)
14. Porter, M.F.: Snowball: A language for stemming algorithms, `http://snowball.tartarus.org/texts/introduction.html`
15. Tokunaga, T., Makoto, I.: Text categorization based on weighted inverse document frequency. Technical report, Tokyo Institute of Technology (1994)
16. Wu, H., Salton, G.: A comparison of search term weighting: Term relevance vs. inverse document frequency. In: Proceedings of the 4th Special Interest Group on Information Retrieval, pp. 30–39 (1981)
17. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information Processing & Management 24(5), 513–523 (1988)
18. Fabbri, F.: Browsing music spaces: Categories and the musical mind (1999)
19. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2(2), 121–167 (1998)
20. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann (2005)
21. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. In: Advances in Kernel Methods, pp. 185–208 (1999)
22. Dhanaraj, R., Logan, B.: Automatic prediction of hit songs. In: Proceedings of International Conference on Music Information Retrieval, pp. 488–491 (2005)
23. Laurier, C., Grivolla, J., Herrera, P.: Multimodal music mood classification using audio and lyrics. In: Proceedings of the 7th International Conference on Machine Learning and Applications, pp. 688–693 (2008)
24. Koppel, M., Schler, J., Argamon, S.: Computational methods in authorship attribution. Journal of the American Society for Information Science and Technology 60(1), 9–26 (2009)
25. HaCohen-Kerner, Y., Beck, H., Yehudai, E., Rosenstein, M., Mughaz, D.: Cuisine: Classification using stylistic feature sets and/or name-based feature sets. Journal of the American Society for Information Science and Technology 61, 1644–1657 (2010)