Impact of Genre in the Prediction of Perceived Emotions in Music

Felipe Souza Tanios[12*], **Tiago Fernandes Tavares**[12†]

[1]Intedisciplinary Nucleus for Sound Studies
University of Campinas
R. da Reitoria, 165, Cidade Universitária – 13083-590 – Campinas - SP

[2]Laboratory of Digital Signal Processing for Telecommunications
School of Electrical and Computer Engineering
University of Campinas Av. Albert Eintein, 400 – 13083-970 – Campinas - SP

first_ftanios@nics.unicamp.br, second_tavares@dca.fee.unicamp.br

Abstract

In on-line music streaming systems and music digital libraries, the emotions that people perceive while listening to a track have become a important criterion while trying to find what to listen to. There are several possible solutions that use emotion as a criterion. Some of them separate and predicts emotions in a track learning individually from its user which demands a considerable amount of time to generate adequate result. Others predict in a generalized way for every user, so it does not take into account musical preferences specifically of its user. People that identify with a social group tend to perceive the same emotions in music and social groups' members often identify with the musical genre. In this paper we describe a method for music emotions prediction using genre information and compare it with a similar classifier that does not use genre information. Results show that prediction accuracy improves in all tested genres, except for one. This suggests that music in different genres convey emotions using different means.

1. Introduction

Online music streaming systems provide a large audio dataset to users. An effective search method is important to deal with such a large amount of data. One possible criterion for music search is the emotion users perceive in music [1].

This work analyzes music as a socio-cultural construct. In this context, music can be seen as a method for communication and social interaction [2]. Furthermore, music genres are often part of the identity of a group that shares common interests [3].

People that belong to the same social group are surrounded by the same social construct. Since emotions are socio-cultural constructs [4], we can reckon that people in the same social group share the similar emotion perceptions.

We use Hanslick's aproach regarding music and emotions [5]. He argues that music itself does not bring an embedded emotion. Instead, it carries a "musical idea", which evokes a perceived emotion on its listener. Perceived emotions are more likely to be used in this type of study because the assessment of "felt" emotion requires personal and situational factors to be considered [6].

There are several proposed models for the description of emotions. The categorical model [7] of Ekman's "Big Six"[8] describes six transculturally perceived emotions (surprise, fear, disgust, anger, happiness and sadness). However, experiments conducted by Livingstone and Brown [9] show that music can consistently communicate happy, sad and dreamy emotions, but this consistency was not found in other emotions.

The automatic prediction of perceived emotion has often relied on mapping pieces to a content-inspired vector space in which close vectors depict similar tracks. This approach was proposed for automatic music genre classification [10] and then adapted to the prediction of other labels, such as emotions [1].

In our work, we evaluated two different struc-

*Supported by PIBIC-UNICAMP.
†Supported by FAPESP

tures for automatic prediction of perceived emotion. In the first, we simply applied audio texture-based classification [10] using perceived emotion as labels. In the second one, we trained a different classifier for each genre in the dataset. This experiment is further discussed in Section 2. The results are shown in Section 3. Finally, Section 4 concludes the paper.

2. Method

The classifier implemented in this work is based on Tzanetakis and Cook's algorithm for music genre classification [10]. The classifier extracts framewise features (spectral centroid, spectral roll-off, spectral flatness, spectral flux, and 30 MFCCs) from a track and calculates their mean and average over a 1s-long sliding window. After that, it calculates the mean and average of the means and averages for each feature within each window. This yields a 136-dimensional vector that maps the track to a content-meaningful $\mathbb{R}^{136}$ vector space. Thus, tracks that are near to each other in this vector space are more similar to each other in terms of auditory texture content. After mapping, the algorithm performs supervised classification within this vector space. In our work, we used the well-known Support Vector Machine (SVM) algorithm for classification.

Our tests aimed at evaluating the impact of using genre information on the prediction of perceived emotion in music. For this purposed, we developed a dataset containing audio files and their respective genre and perceived emotion labels.

To build our dataset we used anonymous discussion websites with graded comments (e.g. Reddit [1]), so we could fetch the average group's emotion perceived for each track. We gathered tracks that were more highly voted in the discussion threads. The first dataset consisted of 200 tracks devided into three emotions, namely Happy, Sad and Dreamy, following previous work by Livingstone and Brown [9].

While analyzing the dataset it became clear to us that the community usually agreed the classification happy and sad tracks but did not agree about dreamy tracks. This left us with a total of only 17 dreamy tracks. Because this is much smaller number than the 90 sad tracks and 93 happy tracks, we decided not to use this emotion in our dataset.

The dataset consist of 183 full-length tracks, divided into five different genres (Indie-Rock, Jazz, Heavy-Metal, Bossa-Nova and Classical music), labeled according to two different perceived emotions. The number of tracks for each genre and emotion label is shown in Table 1.

Genre	Number of tracks		
	Sad	Happy	Total
Jazz	18	19	37
Heavy-Metal	18	19	37
Indie-Rock	18	19	37
Classical	18	17	35
Bossa Nova	18	19	37

Table 1: Number of tracks for each genre and emotion label in the dataset.

We performed two different tests, as shown in Figure 1. The first one uses the whole dataset to train the same classifier, so it would not consider genre information for emotion prediction. The second one trains a different classifier for each genre. All tests were performed using a 10-fold cross validation schema.

The results of both tests are further discussed in the next section.

3. Results and Discussion

The average F1-Score results for Test 1 and Test 2, as defined in Figure 1, are shown in Table 2. As it can be seen, the F1-score in Jazz, Heavy-Metal, Indie-Rock and Classical Music are better than the result related to not taking genre into account. The result for Bossa-Nova, however, is worse than all the others.

We performed a T-Test comparing the results related to each genre-specific classification process and the results related to the classification using all genres at the same time. The result

[1] http://www.reddit.com/

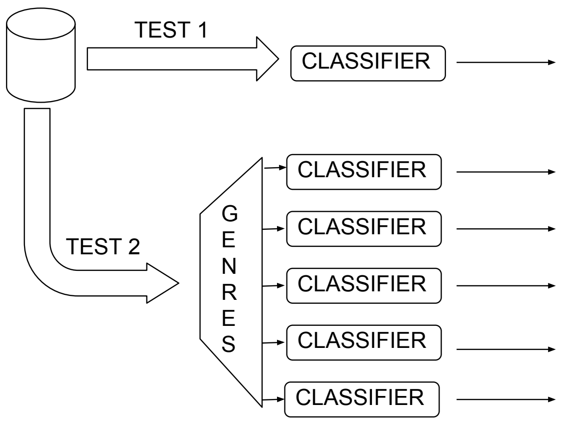

Figure 1: Experiment Flowchart. In test 1, we the dataset containing all genres to train a unique classifier. In test 2, we train a different classifier for each genre.

Test	Genre	F1-score	P
1	All	0.77 ± 0.10	-
2	Jazz	0.85 ± 0.21	0.26
	Heavy-Metal	0.79 ± 0.14	0.11
	Indie-Rock	0.85 ± 0.24	0.76
	Classical	0.86 ± 0.24	0.25
	Bossa Nova	0.56 ± 0.21	0.0005

Table 2: Mean and standard deviations of F1-scores of each genre individually and all genres together. The reported P-Values are related to a t-test between the results related to the classifiers for each genre and the classifier using all genres.

of this test, reported in Table 2, shows that only Bossa-Nova had a statistically significant difference ($P < 0.05$) related to the full set.

This result can be related to the fact that all used features are related to timbre, thus lyrics are ignored. Classical music is commonly built on timbre variations, using instrumentation variation and harmony techniques as compositional tools. Likewise, Heavy-Metal, Indie-Rock and Jazz are musical styles in which frequently there are perceptible timbre differences between happy and sad songs. Bossa-Nova, on the other hand, is a genre in which acoustic guitar and voice are frequently used instruments and the rhythm and intonations are more uniform throughout the style[11]. This means that, although listeners can perceive different emotions in Bossa-Nova song, they cannot be adequately predicted using timbre information.

Therefore, our results indicate that emotion prediction systems should rely on genre information in order to achieve better results. This corroborates with the mixed texture and lyrics methodology used by Yang and Chen [12] and the ideas used as basis by Hu and Downie [1]. However, these results show that improvements on emotion prediction due to the use of both lyrics and texture are related to the genre specificities of each dataset. Also, these results suggest that lyrics and textures should be weighted differently in the prediction of perceived emotion in each genre.

The next section presents conclusive remarks.

4. Conclusion

In this work, we evaluated the impact of using genre information in the prediction of perceived emotion in music. For such, we created a dataset in which tracks were labeled according to their genre and their perceived emotion. Our results show that using a specific classifier for each genre yields better results than using a single classifier, with no genre information, for all genres, except for Bossa-Nova.

This means that musical texture is not an effective feature to explain the perceived emotion in Bossa-Nova. Thus, emotions in different genres are perceived through different means. Therefore, genre information can be used to improve automatic prediction of perceived emotions, which can potentially improve the accuracy of online music suggestion systems. This poses an interesting direction for future work.

References

[1] Xiao Hu and J. Stephen Downie. Improving mood classification in music digital libraries by combining lyrics and audio. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, JCDL

'10, pages 159–168, New York, NY, USA, 2010. ACM.

[2] Yi-Hsuan Yang and Homer H. Chen. Prediction of the distribution of preceived music emotions using discrete samples. *IEEE Transactions on Audio, Speech and Language Processing*, 19(5):2184–2196, 2002.

[3] Beatriz Ilari. Música, comportamento social e relações interpessoais. *Psicologia em Estudo*, 11(1):191–198, 4 2006.

[4] MAURO GUILHERME PINHEIRO KOURY. *Emoções, sociedade e cultura*. EDITORA CRV, 2009.

[5] Eduard Hanslick. *Do Belo Musical*. Editora UNICAMP, 1989.

[6] Alf Gabrielsson. Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, 5(1_suppl):123–147, 2001.

[7] Paula Dornhofer Paro Costa. *Two-Dimensional Expressive Speech Animation - Animação 2D de Fala Expressiva*. PhD thesis, Universidade Estadual de Campinas - Faculdade de Engenharia Elétrica e de Computação, 2 2015.

[8] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, Palo Alto, 1978.

[9] Steven R. Livingstone and Andrew R. Brown. Dynamic response: Real-time adaptation for music emotion. pages 105–111, 2005.

[10] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[11] Bryan McCann. Blues and samba: Another side of bossa nova history. *Luso-Brazilian Review*, 44(2):21–49, 2007.

[12] Dan Yang and WonSook Lee. Disambiguating music emotion using software agents. In *Proceedings of the 5th International Conference on Music Information Retrieval*, Barcelona, Spain, October 10-14 2004.