POLITECNICO
MILANO 1863

Dipartimento di Elettronica, Informazione e Bioingegneria

Master Degree in Computer Science and Engineering

Thesis Title

and its subtitle

by:
Gioele Pozzi

matr.:
10454628

Supervisor:

Co-supervisor:
Clara Borrelli

Academic Year
2019-2020

POLITECNICO
MILANO 1863

Dipartimento di Elettronica, Informazione e Bioingegneria

Master Degree in Computer Science and Engineering

Titolo Tesi

e sottotitolo

Candidato:
Gioele Pozzi

matricola:
10454628

Relatore:

Co-relatore:
Clara Borrelli

Anno Accademico
2019-2020

Abstract

One of the most attractive functions of music is that it can convey emotion and modulate a listener's mood [1]. Music can bring to tears, console us when we are grieving and drive us to love.
Most important thing is that music information behavior studies have identified emotion as an important criterion used by people in music searching and organization. Now become important the field of music emotion recognition.

Sommario

Piacere, so Mario

Acknowledgements

This thesis is the result of almost a year of work at the Image and Sound Processing Lab. First I would thank my supervisor...

Thanks to friends.

Thanks to family.

N.S.

Contents

List of Figures

List of Tables

1
Introduction

1.1 Motivation

Music has an important role in human life. More important, is that music in capable to evoke different emotions for people, but how is structured the relationship between music and emotion? We don't know yet. It's a hard problem, which have very different fields of background, from computer science, machine learning and psychology.

Emotion-aware Music Information Retieval has been difficult due to the subjectivity and temporality of emotion responses to music. The role of physiological signals related to emotions could potentially be exploied in emotion-aware music discovery.

Music is the veichle for emotions, feelings, passion and actions. With the music the composer create a narration which is purely emotional.

Can we measure emotions related to music?

1.2 Outline of the thesis

This thesis is organized as follows:

After a brief intoduction about the objective of the thesis, in chapter 2 and 3 is presented a complete overview about the main arguments in chapter 2, as Music Information Retieval (MIR) and Music Emotion Recognition (MER), Electrodermal Activity (EDA) and other physiological data using on-body sensors.

Chapter 4 is devoted to a complete overview of the state of the art about the main aspects related to chapters 2 and 3 of this thesis, in order to have a general idea about what has been done in the past and which results they have achieved.

In chapter 5 is presented how the dataset we have considered is structured and what results they have reached. Is also shown our implementation of the problem.

Chapter 6 is about the results we have achieved and the comparison between the PMEmo performances.

Finally Chapter 7, draws the conclusions and outlines possible future research directions.

1.3 Application fields

The work proposed in this thesis finds potential application in several fields. Thanks to the work of PMEmo that created a large dataset containing emotion annotations and electrodermal activity signal, we have the possibility to study the relationship between music emotion and physiological signals.
Music Browsing can be an important field of application, because it helps in general in finding, generally in large datasets, what music user are looking for. For example one application could be to create a playlist based on the emotion that songs produce in each of us. Another important application is given by underrstanding the relationship between music and emotion, which is a well known relationship but hard to find structural connection between the two.

2
Theoretical Background on MIR and MER

This chapter introduces the readers to the main basics about Music Information Retrieval and Music Emotion Recognition.

2.1 Music Infomation Retrieval

Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music. MIR is a small but growing field of research with many real-world applications. Those involved in MIR may have a background in musicology, psychoacoustics, psychology, academic music study, signal processing, informatics, machine learning, optical music recognition, computational intelligence or some combination of these.

MIR is being used by businesses and academics to categorize, manipulate and even create music.
A few application to MIR can be:

- Recommender systems: several already exixst, but few are based upon MIR techniques, instead making use of similarity between users or laborius data complilation as in Pandora[1].
- Intelligent and adaptive digital audio effects: aim of design a system that determine the settings of audio effects based on the audio content.

[1]https://www.pandora.com

- Track separation and instrument recognition: like extracting the original tracks as recorded, which could have more than one instrument played per track. Instrument recognition is about identifying the instruments involved into one track.
- Automatic music transcription: process of converting na audio recording into symbolic, such score or a MIDI file.
- Automatic categorization: common task of MIR is musical genre categorization and is the usual task for the yearly Music Information Retrieval Evauation eXchange (MIREX).

2.2 Music Emotion Recognition

Music Emotion Recognition (MER) aim to research on modeling humans emotion perception of music [2], a research topic that emerges in the face of the explosice growth of digital music. Automatic MER allows users to retrive and organiza their music collections in a fashion that is more content-centric than conventional metadata-based methods.
The main challenge is based on the human perception of emotions, their subjective nature of emotion perception. Building such a music emotion recognition system, however, is challenging be- cause of the subjective nature of emotion perception. One needs to deal with issues such as the reliability of ground truth data and the difficulty in evaluating the prediction result, which do not exist in other pattern recognition problems such as face recognition and speech recognition.
MER methods developed try to address the issues related to the ambiguity and granularity of emotion description, the heavy cognitive load of emotion annotation, subjectivity of emotion perception, and the semantic gap between low-level audio signal and high-level emotion perception.

2.2.1 Importance of Music Emotion Recognition

Music plays an important role in human life, even more in the digital age. Never before has such a large collection of music been created and accessed daily by people. Before with the use of compact audio formats with near CD quality such as MP3 and now on with the various streaming services, have greatly contributed to the trmendous growth of digital music libraries.

Conventionally, the management of music collections is based on catalog metadata, such as artist name, album name, and song title. As the amount of content continues to explode, this conventional approach may be no longer sufficient. The way that music information is organized and retrieved has to evolve to meet the ever increasing demand for easy and effective information access.

Music, is a complex acoustic and temporal structure, it is rich in

content and expressivity. When an individual engages with music as a composer, performer or listener, a very board range of mental processes is involved, including *representational* and *evaluative*. The representational process includes the perception of meter, rhythm, tonality, harmony, melody, form, and style, whereas the evaluative process includes the perception of preference, aesthetic experience, mood, and emotion. The term evaluative is used because such processes are typically both valenced and subjective. Both the representational and the evaluative processes of music listening can be leveraged to enhance music retrieval. According to a study of Last.fm[2], emotion tagging is the third most frequent type of tags (first is genre and second locale) assigned to music pieces by online users.
Even if emotion-based music retrieval was a new idea, a survey conducted in 2004 from [3] showed that about 28.2% of the participants identified emotion as an important criterion in music seeking and organization.
The table 2.1 represent the responses of 427 subjects to the question *"When you search for music or music informations, how likely are you to use the following search/browse options?"* [3].

Search/Browse by	Positive rate
Singer/Performer	96.2%
Title of work(s)	91.6%
Some wors of the lyrics	74.0%
Music style/genre	62.7%
Reccomendations	62.2%
Similar artist(s)	59.3%
Similar music	54.2%
Associated usage	41.9%
Singing	34.8%
Theme(main subject)	33.4%
Popularity	31.0%
Mood/emotional state	**28.2%**
Time period	23.8%
Occasions to use	23.6%
Instrument(s)	20.8%
Place/event where heard	20.7%
Stroryline of music	17.9%
Tempo	14.2%
Record label	11.7%
Publisher	6.0%

Table 2.1: Responses of 427 subjects to the question *"When you search for music or music informations, how likely are you to use the following search/browse options?"*

[2]https://www.last.fm/home

Into another survey [4], they present findings from an exploratory questionnaire study featuring 141 music listeners (between 17 and 74 years of age) that offers some novel insights.
One of the most exciting but difficult endeavors in research on music is to understand how listeners respond to music. It has often been suggested that a great deal of the attraction of music comes from its "emotional powers". That is, people tend to value music because it expresses and induces emotions. The table 2.2 tries to resume the motivations to the answer *"Why do we listen to music?"*

Motive	Ratio
"To express, release and influence emotions"	47%
"To relax and settle down"	33%
"For enjoyment, fun, and pleasure"	22%
"As company and background sound"	16%
"Because it makes me feel good"	13%
"Because it's a basic need, I can't live without it"	12%
"Because I like, love music"	11%
"To get energized"	9%
"To evoke memories"	4%

Table 2.2: Responses of 141 subjects to the question *"Why do you listen to music?"*

Some music companies, like Allmusic.com[3], gives the possibility to search music by emotion labels. With these, the user can retrive and browse artists or albums by emotion.

Making computers capable of recognizing the emotion of music also enhances the way humans and computers interact. It is possible to play back music that matches the user's mood detected from physiological, prosodic, or facial cues. A cellular phone equipped with automatic music emotion recognition (MER) function can then play a song best suited to the emotional state of the user; a smart space (e.g., restaurant, conference room, residence) can play background music best suited the people inside it.

2.2.2 Recognizing the perceived emotion of music

There is a relationship between music and emotions, that has been the subject of much discussion and research in many different disciplines, like philosphy, musicology, sociology.
In psychological studies, emotion are often divided into three categories:

- *Expressed emotion*: the ones the performer tries to communicate with the listener.

[3]https://www.allmusic.com/moods

- ***Perceived emotion***: represented by music and perceived by the listener.
- *Felt or Evoked emotion*: induced by music and felt by the listener.

MER focus on perceived emotions because they are less subjective than felt emotions and are often easier to conceptualize. This because felt emotions depends on personal factors and the situation in which the listener processes the song. From an engineering point of view, one of the main interests is to develpo a computational model of music emotion and to facilitate emotion-based music retrieval and organization. MIR community has made many efforts for automatic recogition of the perceived emotion of music, various implementations will be presented further in chapter 4.
A typical approach to MER categorizes emotions into a number of classes and applies Machine Learning (ML) techniques to train a classifies. Usually are extracted some features of music to represent the acoustic property of a music piece. Typically, a subjective test is conducted to collect the ground truth needed for training the computational model of emotion prediction. Subjects are asked to report their emotion perceptions of the music pieces.
To learn the relationship between music features and emotion labels have been applied, such as Support Vector Machines (SVMs), Gaussian Mixture Models (GMMs), Neural Networks (NN) and k-nearest neighbor.
After training, the automatic model can be applied to classify the emotion of an input music piece, for example a schematic diagram of the *cathegorical approach* to MER can be seen in figure 2.1.

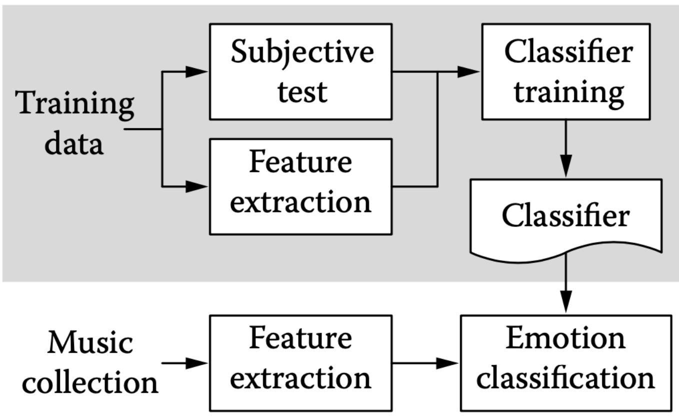

Figure 2.1: Schematic diagram of the categorical approach to MER

2.2.3 Open issues of Music Emotion Recognition

As MER is a quite new domain, there are some elements that have no clear answer. Four of these issues are:

1. Ambiguity and Granularity of emotion description:

2. Heavy cognitive load of emotion annotation

3. Subjectivity of emotional perception:

4. Semantic gap between Low-Level (LL) and audio signal and High Level (HL) Human perception:

3
Theoretical Background on EDA

This chapter introduces the readers to...

3.1 Some different sections

3.2 Remarks

4
State of the Art

This chapter introduces the models

4.1 Some different sections

4.2 Conclusive Remarks

In this chapter we introduce the main issues ...

5
Implementation and Results

In this chapter we present...

Then...

Finally ...

5.1 Some different sections

5.2 Conclusive Remarks

6
Dataset Improvements

In this chapter we present...

Then...

Finally ...

6.1 Some different sections

6.2 Conclusive Remarks

7
Conclusions and Future Works

This work of thesis proposes a methodology for...

The devised methodology is based on...

The main advantages are...

As far as the experiments are concerned...

The proposed approach has shown promising results both in simulation and in the experiments.

7.1 Future Works

Generalization We would like to generalize...

Challenging scenarios Another possible improvement is related to the extension of the proposed approach to...

Different approaches Finally we are moving towards a deeper analysis of... a a a a a a a a a a a a a a a a

Appendices

A Equipment 1

B Proofs of Mathematical Theories1

Bibliography

[1] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, (Hangzhou, China), pp. 375–376, 2003.

[2] Y.-H. Yang and H. H. Chen, *Music emotion recognition.* USA: CRC Press, Inc., 1st ed., 2011.

[3] J. H. Lee and J. S. Downie, "Survey of music information needs, uses, and seeking behaviours: preliminary findings.," in *ISMIR*, vol. 2004, p. 5th, Citeseer, 2004.

[4] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal of new music research*, vol. 33, no. 3, pp. 217–238, 2004.