

Studies in Computational Intelligence

Volume 747

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/7092>

Jacek Grekow

From Content-based Music Emotion Recognition to Emotion Maps of Musical Pieces

Jacek Grekow
Faculty of Computer Science
Bialystok University of Technology
Bialystok
Poland

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-3-319-70608-5 ISBN 978-3-319-70609-2 (eBook)
<https://doi.org/10.1007/978-3-319-70609-2>

Library of Congress Control Number: 2017957686

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my great Family

Foreword

More and more multimedia systems use emotions as a factor in their practical applications. They try to recognize emotions in all forms of content in which humans are emotionally involved—in other words, in texts, music, or films. An example of such solutions is giant music libraries available via the Internet with music search systems using various criteria. Emotions turned out to be one of the more attractive and novel criteria for searching them. Just searching for compositions by title, composer and/or genre was not enough, and adding emotions as an option improved the attractiveness and usefulness of the search systems.

To be able to index music files in terms of emotions, they first have to be identified. Music emotion recognition (MER) is an interdisciplinary problem that affects such fields as signal processing, machine learning, psychology, music theory, and practice. MER requires researchers to be familiar with various distant fields, which on one hand is a complex task and on the other tremendously exciting.

This book presents particular issues that pertain to the creation of a music emotion detection system, such as: deciding on an emotion model, creating ground truth data, extracting and selecting features, and constructing a prediction model. The author focused on content-based MER, i.e., on analyzing the structures contained in music files and finding a relationship with emotions. He focused on examining audio as well as MIDI files and for each presented the relevant feature sets that describe them.

In this work, MER is presented as a classification and regression problem, which is closely connected with the selected emotion model: categorical and dimensional, respectively. Taking into consideration the way humans perceive emotions, the author decided to focus on perceived emotions in music and not felt emotions. During training data construction, he takes advantage of musicians' knowledge who expressed their opinions on the emotions they observed within musical excerpts. The author takes the reader through the specific stages of the emotion recognition system construction and presents the path from music files to the emotion maps created from them, which visualize emotion distribution over time. The book also puts forth practical applications of automatic emotion detection that were tested experimentally.

Emotions in music are very much connected with humans; we could say that they have a very human nature. They accompany man while composing, performing, or listening to music. Let us hope that computer systems evaluating emotions will not replace humans in creating them, but will only aid us in better understanding their relationship with music.

This book can be seen as a source of information for scientists, academicians, teachers, and students involved in constructing automatic emotion detection systems, as well as musicologists studying and analyzing musical compositions. Additionally, it can serve as an inspiration for other similar experiments conducted with the aim of analyzing emotion distribution in music.

Zbigniew W. Ras
Professor & KDD Lab Director
University of North Carolina Charlotte, USA
Professor, Polish-Japanese Academy of Information
Technology PJAIT
Warsaw, Poland

Preface

This book presents the most important issues with automated systems for music emotion recognition. These problems include emotion representation, annotation of music excerpts, feature extraction, and machine learning. The book concentrates on presenting content-based analysis of music files, which automatically analyzes the structures of a music file and annotates this file with the perceived emotions. Emotion detection in MIDI and audio files is presented.

In the experiments, the categorical and dimensional approaches were used, while for music file annotation, the knowledge and expertise of music experts with a university music education. The built automatic emotion detection systems enable the indexing and subsequent searching of music databases according to emotion. The obtained emotion maps of musical compositions provide new knowledge about the distribution of emotions in music and can be used to compare the distribution of emotions in different compositions as well as for emotional comparison of different interpretations of one composition.

Białystok, Poland
May 2017

Jacek Grekow

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Organization of This Book	2
Part I Emotion in Music		
2	Representations of Emotions	7
2.1	Perception of Emotions	7
2.2	Categorical Approach	7
2.3	Dimensional Approach	10
2.4	Summary	11
3	Human Annotation	13
3.1	Introduction	13
3.2	Length of a Musical Segment	13
3.3	Audio Music Data	15
3.3.1	Data Set	15
3.3.2	Music Experts	16
3.3.3	Annotation Process	16
3.3.4	Results	17
3.4	MIDI Music Data	20
3.4.1	Data Set	20
3.4.2	Hierarchical Emotion Model	21
3.4.3	Annotation Process	22
3.4.4	Results	22
3.5	Summary	23
Part II Emotion Detection in MIDI Files		
4	MIDI Features	27
4.1	Introduction	27

4.2	MIDI Features Tools	28
4.3	Description of MIDI Features	28
4.3.1	Rhythm Features	29
4.3.2	Harmony Features	33
4.3.3	Harmony-Rhythm Features	38
4.3.4	Dynamic Features	40
4.4	Conclusions	40
5	Hierarchical Emotion Detection in MIDI Files	43
5.1	Introduction	43
5.2	Related Work	43
5.3	MIDI Music Data	44
5.4	Feature Extraction	45
5.5	Construction of Classifiers	46
5.5.1	First Level Classifiers	46
5.5.2	Second Level Classifiers	49
5.6	Hierarchical Classification	50
5.7	Emotion Tracking in MIDI Files	51
5.7.1	System Construction	51
5.7.2	Musical Composition Segmentation	52
5.8	Results of Emotion Tracking in MIDI Files	53
5.8.1	Emotion Histograms of Musical Compositions	53
5.8.2	Emotion Maps	54
5.8.3	Quantity of Changes of Emotion	56
5.9	Conclusions	57
 Part III Emotion Detection in Audio Files		
6	Audio Features	61
6.1	Introduction	61
6.2	Features from Audio Analysis Tools	61
6.2.1	Essentia	62
6.2.2	Marsyas	63
6.3	Selected Audio Features and Labeled Emotions	63
6.3.1	Timbre Features	64
6.3.2	Rhythm Features	70
6.3.3	Tonal Features	71
6.4	Conclusions	74
7	Detection of Four Basic Emotions	75
7.1	Introduction	75
7.2	Related Work	75
7.3	Music Data	77
7.4	Feature Extraction	78
7.5	Results	78

7.5.1	Construction of One Classifier Recognizing Four Emotions	78
7.5.2	Construction of Binary Classifiers	79
7.5.3	Evaluation of Different Combinations of Feature Sets	81
7.5.4	Emotion Maps	81
7.6	Conclusions	83
8	Emotion Tracking of Radio Station Broadcasts	85
8.1	Introduction	85
8.2	System Construction	86
8.3	Music Data	87
8.3.1	Training Data	87
8.3.2	Recorded Radio Broadcasts	87
8.4	Feature Extraction	88
8.5	Construction of Classifiers	88
8.5.1	Music/Speech Classifier	88
8.5.2	Classifier for Emotion Detection	89
8.6	Results of Emotion Tracking of Radio Stations	89
8.6.1	Comparison of Radio Stations	90
8.6.2	Emotion Maps of Radio Station Broadcasts	90
8.7	Conclusions	93
9	Music Emotion Maps in the Arousal-Valence Space	95
9.1	Introduction	95
9.2	Related Work	95
9.3	Music Data	96
9.4	Feature Extraction	97
9.5	Regressor Training	97
9.6	Evaluation of Different Combinations of Feature Sets	98
9.7	Selected Features Dedicated to the Detection of Arousal and Valence	99
9.8	Emotion Maps	101
9.8.1	Emotion Maps of Two Compositions	101
9.8.2	Features Describing Emotion Maps	103
9.8.3	Comparison of Musical Compositions	103
9.9	Conclusions	106
10	Comparative Analysis of Musical Performances by Using Emotion Tracking on the Arousal-Valence Plane	107
10.1	Introduction	107
10.2	Related Work	107
10.3	System Construction	108
10.4	Music Data for Regressor Training	108
10.5	Regressor Training	110

10.6	Aligning of Audio Recordings	110
10.7	Analyzed Performances	112
10.8	Results	113
10.8.1	Influence of Recording Alignment on Emotion Values of the Compared Segments	113
10.8.2	Performances and Arousal over Time	115
10.8.3	Performances and Valence over Time	116
10.8.4	Performances and Arousal-Valence over Time	117
10.8.5	Arousal-Valence Trajectory	119
10.8.6	Scape Plot	119
10.8.7	Arousalcape	121
10.8.8	Valencescape	122
10.8.9	AVscape	123
10.9	Evaluation	124
10.9.1	Parameters Describing the Most Similar Performances	124
10.9.2	Ground Truth for Performing Similarity Assessments Between Performances	125
10.9.3	Evaluation Parameters	126
10.10	Conclusions	129
	References	131
	Index	137