

DOCTORAL THESIS

Emotion-based music retrieval and recommendation

Deng, Jie

Date of Award:
2014

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

The digital music industry has expanded dramatically during the past decades, which results in the generation of enormous amounts of music data. Along with the Internet, the growing volume of quantitative data about users (e.g., users' behaviors and preferences) can be easily collected nowadays. All these factors have the potential to produce big data in the music industry. By utilizing big data analysis of music related data, music can be better semantically understood (e.g., genres and emotions), and the user's high-level needs such as automatic recognition and annotation can be satisfied. For example, many commercial music companies such as Pandora, Spotify, and Last.fm have already attempted to use big data and machine learning related techniques to drastically alter music search and discovery. According to musicology and psychology theories, music can reflect our heart and soul, while emotion is the core component of music that expresses the complex and conscious experience. However, there is insufficient research in this field. Consequently, due to the impact of emotion conveyed by music, retrieval and discovery of useful music information at the emotion level from big music data are extremely important.

Over the past decades, researchers have made great strides in automated systems for music retrieval and recommendation. Music is a temporal art, involving specific emotion expression. But while it is easy for human beings to recognize emotions expressed by music, it is still a challenge for automated systems to recognize them. Although some significant emotion models (e.g., Hevner's adjective circle, Arousal-

Valence model, Pleasure-Arousal-Dominance model) established upon the discrete emotion theory and dimensional emotion theory have been widely adopted in the field of emotion research, they still suffer from limitations due to the scalability and specificity in music domain. As a result, the effectiveness and availability of music retrieval and recommendation at the emotion level are still unsatisfactory.

This thesis makes contribution at theoretical, technical, and empirical level. First of all, a hybrid musical emotion model named “Resonance-Arousal-Valence (RAV)” is proposed and well constructed at the beginning. It explores the computational and time-varying expressions of musical emotions. Furthermore, dependent on the RAV musical emotion model, a joint emotion space model (JESM) combines musical audio features and emotion tags feature is constructed. Second, corresponding to static musical emotion representation and time-varying musical emotion representation, two methods of music retrieval at the emotion level are designed: (1) a unified framework for music retrieval in joint emotion space; (2) dynamic time warping (DTW) for music retrieval by using time-varying music emotions. Furthermore, automatic music emotion annotation and segmentation are naturally conducted. Third, following the theory of affective computing (e.g., emotion intensity decay, and emotion state transition), an intelligent affective system for music recommendation is designed, where conditional random fields (CRF) is applied to predict the listener’s dynamic emotion state based on his or her personal historical music listening list in a session. Finally, the experiment dataset is well created and proposed systems are also implemented. Empirical results (recognition, retrieval, and recommendation) regarding accuracy compared to previous techniques are also presented, which demonstrates that the proposed methods enable an advanced degree of effectiveness of emotion-based music retrieval and recommendation.

Keywords: Music and emotion, Music information retrieval, Music emotion recog-

dition, Annotation and retrieval, Music recommendation, Affective computing, Time series analysis, Acoustic features, Ranking, Multi-objective optimization

Acknowledgements

First and foremost I would like to record gratitude to my principal supervisor, Prof. **Clement Leung**, for giving me the opportunity to work with his supervision. His feedback often gave very insightful perspectives on the aspects to the research. More importantly, the monitoring I have received from him extends well beyond academic research. I am thankful for his continued guidance and encouragement.

Next, I would like to appreciate Dr. **Chen Li** for serving as my co-supervisor. She has also provided insightful discussions about the research. In addition, many thanks to my research committee, Dr. **Chu, Xiao Wen** and Dr. **Chun Hung Li**, who spent time out of their busy schedules to provide valuable feedbacks and their constructive suggestions on my research work in PhD studies.

Then, I would like to thank my thesis committee members, Prof. **Ricky N. S. Wong**, Prof. **Qing Li**, Prof. **Jian Liang Xu**, Dr. **Chu, Xiao Wen**, Prof. **Yiu Wing Leung**, Prof. **Pong Chi Yuen**. I want to thank them for letting my defense be an enjoyable moment, and for their brilliant comments and suggestions.

It is also important to recognize the unconditional support from the people at the department of computer science, Hong Kong Baptist University, specially **Kristin Li**, and **Andy Chen**. I would like to thank all my friends and especially those supported me all the time through my research work. Furthermore, I would especially like to thank Italian professor **Alfredo Milani** and friends Dr. **Valentino Santucci** and **Valentina Franzoni**.

Last but definitely not least, I especially thank my parents **Zongyin Deng, Xiaolan Leng**, and my brother **Kai Deng** for their unconditional love, care, support, and encouragement. I love them so much, and I would not have made it this far without them. Thank you all so much.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	v
Table of Contents	vii
List of Tables	xiii
List of Figures	xvi
Chapter 1 Introduction	1
1.1 Big Data in Music Industry	1
1.2 Motivation and Significance	2
1.3 Objectives of the Thesis	3
1.4 Organization of the Thesis	4
Chapter 2 Literature Review	6
2.1 Music and Emotion: Theory and Research	7
2.1.1 Music Concept and Terminology	7
2.1.2 Music and Emotion	10
2.2 Emotion Theories	11

2.2.1	Discrete Emotion Theory	12
2.2.2	Dimensional Emotion Theory	15
2.2.3	Comparison of the Discrete and Dimensional Emotion Models	18
2.3	Audio Feature Extraction and Selection	19
2.3.1	Physical Features	19
2.3.2	Statistical Features	19
2.3.3	Perceptual Features	21
2.3.4	Semantic Features	21
2.3.5	Feature Selection	22
2.4	Music Emotion Recognition	24
2.4.1	Regression Model	24
2.4.2	Gaussian Mixture Model	25
2.4.3	Hidden Markov Model	25
2.4.4	Miscellaneous	26
2.5	Music Retrieval System	26
2.5.1	Music Retrieval Methods: State of the Art	27
2.6	Music Recommendation System	30
2.6.1	Problem Formulation	30
2.6.2	Music Recommendation Methods: State of the Art	30
2.7	Summary	38

Chapter 3 A Novel Musical Emotion Representation for Music Emotion Recognition 39

3.1	Introduction	40
3.2	Musical Emotion Representation	41
3.2.1	Resonance-Arousal-Valence (RAV) Model	41
3.2.2	Static representation of musical emotion	42

3.2.3	Dynamic representation of musical emotion	44
3.3	Emotion-relevant Feature Extraction	45
3.3.1	Arousal-based Features	46
3.3.2	Valence-based Features	47
3.3.3	Resonance-based Features	48
3.3.4	Feature Selection	48
3.4	Regression Models for Musical Emotion	51
3.4.1	Support Vector Regression (SVR)	51
3.4.2	Sparse Bayesian Regression (SBR)	53
3.4.3	Variational Bayesian Regression (VBR)	55
3.5	Experiment: Music Emotion Recognition	56
3.5.1	Method	56
3.5.2	Evaluation	64
3.5.3	Experimental Results	65
3.6	Summary	69

Chapter 4 Music Retrieval in Joint Emotion Space Combined Audio

	Features and Emotion Tags	71
4.1	Introduction	72
4.2	The Overview of a Unified Framework for Music Retrieval	73
4.3	Joint Emotion Space Model	74
4.3.1	Musical Audio Feature Space	74
4.3.2	Emotion Tag Feature Space	76
4.3.3	Model Selection	77
4.4	Music Emotion Annotation	78
4.4.1	Problem Formulation	78
4.4.2	Supervised Annotation Approach	79

4.5	Music Retrieval in Joint Emotion Space	79
4.5.1	Querying the model	80
4.5.2	Music Emotion Ranking	82
4.6	Experiments	85
4.6.1	Dataset and Experimental Setup	86
4.6.2	Performance Evaluation	86
4.6.3	Experimental Results	88
4.7	Summary	96

Chapter 5 Dynamic Time Warping for Music Retrieval Using Time

Series Analysis of Musical Emotion 99

5.1	Introduction	100
5.2	Time Series Analysis for Musical Emotion	101
5.2.1	Modeling Musical Emotion Using Dynamic Texture	101
5.2.2	Discrete Time Kalman Filtering	103
5.2.3	Discrete Time Kalman Smoothing	105
5.2.4	Parameter Estimation for DT Model	107
5.2.5	Multiple Dynamic Textures (MDT) Model	110
5.2.6	Model Selection for MDT	110
5.3	Emotion-based Music Segmentation Using MDT	111
5.3.1	Problem Formulation	112
5.3.2	Segmentation Method	113
5.4	Matching and Retrieval of Time-Varying Musical Emotion Sequences	113
5.4.1	Dynamic Time Warping for Music Matching	114
5.4.2	Music Retrieval Based on Time-Varying Emotions	115
5.5	Data Collection and Evaluation	116
5.6	Experiments and Results	119

5.7	Summary	126
Chapter 6 Towards Intelligent Affective Recommendation of Music		130
6.1	Introduction	131
6.2	The Overview of Proposed Method for Affective Recommendation . . .	132
6.3	Computational Models for Musical Emotion in Affective System . . .	135
6.3.1	Musical Emotion Intensity	136
6.3.2	Emotion Intensity Decay Model	136
6.3.3	Emotion State Transition Model	138
6.4	Dynamic Emotion-Aware System Modeling	140
6.4.1	Problem Formulation	140
6.4.2	Methodology	140
6.5	The Algorithm of Intelligent Affective Music Recommendation	143
6.5.1	Music Emotion Similarity	143
6.5.2	Music Emotion Ranking	144
6.6	Improving Recommendation Quality	146
6.6.1	Reinforcement Learning: Learning from Interaction	146
6.7	Experiment 1: Dynamic Emotion State Prediction by CRF	151
6.7.1	Method	151
6.7.2	Evaluation Criteria	156
6.7.3	Experimental Results	158
6.8	Experiment 2: Dynamic Affective Recommendation of Music	163
6.8.1	Evaluation Criteria	164
6.8.2	Results	164
6.9	Summary	169
Chapter 7 Summary and Conclusion		173
7.1	Overall Summary	174

7.2 Contributions	177
7.3 Future Work	178
7.4 Conclusions	180
Bibliography	186
Curriculum Vitae	215