DOCTORAL THESIS

Emotion-based music retrieval and recommendation

Deng, Jie

Date of Award:
2014

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-

nition, 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