Dimensional Music Emotion Recognition by Machine Learning

Junjie Bai, School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing, China & School of Instrument Science and Engineering, Southeast University, Nanjing, China
Lixiao Feng, School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing, China
Jun Peng, School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing, China
Junliang Shi, School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing, China
Kan Luo, School of Information Science and Engineering, Fujian University of Technology, Fuzhou, China
Zuojin Li, School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing, China
Lu Liao, School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing, China
Yingxu Wang, Schulich School of Engineering and Hotchkiss Brain Institute, University of Calgary, Calgary, Canada

ABSTRACT

Music emotion recognition (MER) is a challenging field of studies that has been addressed in multiple disciplines such as cognitive science, physiology, psychology, musicology, and arts. In this paper, music emotions are modeled as a set of continuous variables composed of valence and arousal (VA) values based on the Valence-Arousal model. MER is formulated as a regression problem where 548 dimensions of music features were extracted and selected. A wide range of methods including multivariate adaptive regression spline, support vector regression (SVR), radial basis function, random forest regression (RFR), and regression neural networks are adopted to recognize music emotions. Experimental results show that these regression algorithms have led to good regression effect for MER. The optimal $R^2$ statistics and VA values are 29.3% and 62.5%, respectively, which are obtained by the RFR and SVR algorithms in the relief feature space.

KEYWORDS

Emotion Regression, Feature Extraction, Machine Learning, Music Emotion Recognition, Pattern Recognition, Valence Arousal Model

1. INTRODUCTION

Music is not only a form of art but also a language that expresses human emotions, inner modes, and affective information (Juslin, 2001; Rodriguez, Ramos, & Wang, 2012; Wilson, and Keil, 2001; Juslin, & Sloboda, 2001). It is generally believed that music cannot be composed, performed, or comprehended without affective cognition and involvement. Music expresses affective emotions including joy, happiness, annoyance, sadness, pain, etc. Aesthetics and cognitive science recognize music as an effective form of affective expression. However, music experience is a subjective behavior in the process of music creation or appreciation. Individuals may have different understanding of the
same piece of music and different extent of affective emotional effects. Therefore, how the affective emotions of music are formally evaluated is a challenging problem in fields of musicology, esthetics, psychologists, and cognitive science (Juslin, 2001; Hallam, & Thaut, 2008). At present, machine learning algorithm is the research hotspot (Wang, 2016, 2015, 2015), and the machine learning algorithms are widely adapted to the recognize music emotions (Yang et al., 2008; Bang et al., 2013; Mokhsin et al., 2014; Jens, and Sand et al., 2015; Chin, and Lin et al, 2013).

In 1980s, Russell and Thayer proposed the Valence - Arousal model for music emotion description which is widely accepted and used by musicologists, estheticians and psychologists (Russell, 1980). A 2D emotion plane is introduced in the dimensions of valence and arousal (VA). In AV plane, the horizontal axis is defined as valence values representing a positive or negative emotion. The vertical coordinate is defined as arousal values of exciting or calming. Both VA values are ranged in (-1, 1). In this measurement scope, a valence value closer to 1 means a higher and positive emotions, and vice versa. Similarly, a higher arousal value indicates a stronger emotional intensity, and vice versa.

For example, as shown in Figure 1, a happy feeling is an emotion of positive valence and highly arousal, while sad is an emotion of negative valence and low arousal. Therefore, any form of music emotion can be mapped to the AV plane as a certain point. This allows music emotions to be formally recognized by a pair of VA values (Thayer, 1989). This dimensional conceptualization of music emotions provides a simple, reliable, and understandable model for practical affective experiments and manipulations.

Figure 1 provides four quadrants in the music emotion plane corresponding to the categories of positive and strong (1), negative and strong (2), negative and weak (3), as well as positive and weak (4) as described in Table 1.

A music emotion database known as MediaEval (http://www.multimediaeval.org/) is adopted in this work for music emotion recognition. The MediaEval dataset includes 1000 MP3 music archive of songs (http://freemusicarchive.org/), where the emotional features of each song are annotated. Due to redundancy in the database, only 744 annotated songs are identified, which are classified into two categories for learning algorithm training (619 songs) and 125 songs for testing (125 songs), respectively. The length of each song in the database is 45s and the sampling rate is 44100Hz. The annotation file contains a title, a set of 2Hz dynamic VA values, and the standard deviation information for each song (Soleymani et al., 2013).

Figure 1. The valance-arousal plane of music emotions
Bringing the Arts as Data to Visualize How Knowledge Works
www.igi-global.com/chapter/bringing-the-arts-as-data-to-visualize-how-knowledge-works/127496?camid=4v1a

Feature Extraction of Video Using Artificial Neural Network
www.igi-global.com/article/feature-extraction-of-video-using-artificial-neural-network/179932?camid=4v1a