Music Retrieval and Recommendation Scheme Based on Varying Mood Sequences

Sanghoon Jun, Korea University, South Korea
Seungmin Rho, Korea University, South Korea
Eenjun Hwang, Korea University, South Korea

ABSTRACT

A typical music clip consists of one or more segments with different moods and such mood information could be a crucial clue for determining the similarity between music clips. One representative mood has been selected for music clip for retrieval, recommendation or classification purposes, which often gives unsatisfactory result. In this paper, the authors propose a new music retrieval and recommendation scheme based on the mood sequence of music clips. The authors first divide each music clip into segments through beat structure analysis, then, apply the k-medoids clustering algorithm for grouping all the segments into clusters with similar features. By assigning a unique mood symbol for each cluster, one can transform each music clip into a musical mood sequence. For music retrieval, the authors use the Smith-Waterman (SW) algorithm to measure the similarity between mood sequences. However, for music recommendation, user preferences are retrieved from a recent music playlist or user interaction through the interface, which generates a music recommendation list based on the mood sequence similarity. The authors demonstrate that the proposed scheme achieves excellent performance in terms of retrieval accuracy and user satisfaction in music recommendation.

Keywords: Artificial Neural Network, Mood Sequence, Music Recommendation, Music Retrieval, Smith-Waterman Algorithm

1. INTRODUCTION

With the explosive growth of digital media population, content-based retrieval has come to play an important role in the multimedia domain, due to its rich and intuitive expressiveness in query formulation. The details of content-based multimedia retrieval are deeply dependent on the media type. For instance, in the case of audio data, low-level acoustic features such as MFCCs and ZCR are extracted from audio clips and indexed for retrieval purposes (Ghias, Logan, Chanberlin, & Smith, 1995; Birmingham, Dannenberg, & Pardo, 2006; Rho, Han, Hwang, & Kim, 2008). However, these low-level features fail to provide semantic information of music contents such as mood or emotion. This is a serious limitation in retrieving and recommending appropriate music in
many applications. To handle this limitation, many researchers have tried to utilize high-level musical features such as harmonics and beat.

On the other hand, many researchers (Feng, Zhuang, & Pan, 2003; Juslin & Sloboda, 2001; Lu, Liu, & Zhang, 2006) have investigated the influence of music features such as loudness and tonality on the perceived emotional expression. They analyzed such data using diverse techniques, some of which are involved in measuring psychological and physiological correlations between the state of a particular musical factor and emotion evocation. Feng et al. (2003) approached the problem from the viewpoint of Computational Media Aesthetics (CMA) and mapped two dimensions of tempo and articulation into four categories of moods: happiness, anger, sadness and fear. This categorization is based on Juslin’s theory (Juslin & Sloboda, 2001) in which the authors investigated the utilization of acoustic cues in the communication of music emotions between performers and listeners, and measured the correlations between emotional expressions (i.e., anger, sadness and happiness) and acoustic cues (i.e., tempo, spectrum and articulation). Lu et al. (2006) classified various features into three categories: intensity, timber and rhythm and mapped all the moods into Thayer’s two-dimensional space (Juslin & Sloboda, 2001). They used the Gaussian mixture model (GMM) as a classifier. To track music moods, they considered musical mood variation and proposed a mood boundary detection scheme based on threshold adaptation. To the best of our knowledge, no effort has been made to measure the similarities of music clips based on the musical mood change pattern.

Some music genres such as classical usually invoke more than one musical mood. For instance, Beethoven’s Symphony No. 5 is famous for mood changes ranging from magnificence to nimbleness. Distinct musical features invoke different musical moods. In Jun et al. (2009), we considered a single global feature for mood classification and used it for music retrieval. However, in many cases, a single global feature could not summarize various moods in a music clip effectively and hence provide incorrect result. To solve this problem, in this paper, we first divide music clips into segments, extract various low-level features from each segment and then classify them into clusters with similar features. By assigning an appropriate mood symbol for each cluster, we can get musical mood sequences for music clips and use them for music retrieval and recommendation.

The rest of this paper is organized as follows: In Section 2, we present a brief overview on the recent music retrieval and mood classification techniques. Section 3 presents the overall system architecture and some details on the similar music retrieval and recommendation. Section 4 describes the experiments we performed and some of the results. In the last section, we conclude the paper with directions for future work.

2. RELATED WORK

In this section, we introduce some of the recent efforts in the area of music mood recognition, retrieval and recommendation. We first investigate the state of the art musical mood/emotion recognition techniques.

2.1. Musical Mood/Emotion Recognition

Music is a language of emotions, and hence music emotion could play an important role in the various music-related applications such as music understanding, retrieval, and recommendation. Many issues for music emotion recognition have been addressed by different disciplines such as physiology, psychology, cognitive science and musicology. A regression approach for MER (Music Emotion Recognition) was proposed by Yang et al. (2008), where MER was formulated as a regression problem to predict the arousal and valence values (AV values) of each music sample which correspond to a point in the AV plane. As a result, users can efficiently retrieve desired music samples by specifying a point in the emotion plane. For performance improvement, they reduced
Using a Dialogue Manager to Improve Semantic Web Search
www.igi-global.com/article/using-a-dialogue-manager-to-improve-semantic-web-search/149203?camid=4v1a

The Development of an Optimised Metadata Application Profile
www.igi-global.com/chapter/the-development-of-an-optimised-metadata-application-profile/175865?camid=4v1a