http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (18): 4472-4475, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4472.4475

Emotion-Based Music Classification

[1,2]Weina Zhu and [3]Jan Drewes
[1]School of Information Science, Yunnan University, China
[2]Kunming Institute of Zoology, Chinese Acad. of Science, China
[3]Center for Mind/Brain Sciences (CIMeC), Trento University, Rovereto, Italy

Abstract: In traditional content-based music classification or retrieval, classic musical features such as pitch, tempo, melody, timbre and rhythm were used. On the other hand, music conveys and evokes emotions which is one of the most important features of music. However, only limited success has been obtained in studying automatic classifiers of emotion in music. In this study, we constructed a multi-feature-based representation of music emotion by means of combining several musical features. And with this representation, we used a linear classifier to determine the kind of emotion of a certain piece of music. We evaluated the classification on music pieces which were labeled according to the music plane by human subjects.

Key words: Emotion, music classification, features

INTRODUCTION

Music has played a very important role in man's life since the beginning of civilization. With the increasing use of computers and mobile devices, the size of the "global music database" is growing and it is a big challenge to organize and retrieve this content effectively. The traditional text-based retrieval (such as titles, composers) does not meet the requirements of search and classification in massive music databases and content-based retrieval is a desirable method for the challenge of music database retrieval.

In many previous studies on content-based music classification, classic musical features were considered as candidate features, for example: Pitch, tempo, melody, timbre and rhythm (Ghias *et al.*, 1995; McNab and Smith, 1996; Rodger *et al.*, 1996a, b; Naoko *et al.*, 2000; Sonoda and Muraoka, 2000; Won *et al.*, 2005). By using these music features, there have been many improvements in retrieval systems. On the other hand, in everyday life, the reason why people listen to and enjoy music is not to analyze the features of music but to feel the emotions induced by music. Emotional experience is one of the most important reasons why people listen to music. Furthermore, it has been proven that emotion is one of most important higher level features of music (Huron, 2000). However, emotions conveyed by music are not being sufficiently considered in current music retrieval systems. A grasp of emotions in music might be a great help to effectively classify or retrieve music.

In this article, we endeavored to answer the question: Is it possible to develop an algorithm to enable computers to "perceive" emotion of music in a similar way as humans do? On the one hand, when normal humans (perhaps not trained musicians) are listening to music, they do not generally perceive music pitch, melody, timbre, tempo or rhythm separately. On the other hand, all of these basic music features give service to the expression of musical emotion (Ye and Jiang, 1988) which has been showed by cognitive neuroscience studies (Kaminska and Woolf, 2000; Dalla Bella *et al.*, 2001; Geringer and Madsen, 2003; Hailstone *et al.*, 2009). Thus we will construct a multi-feature-based representation of music emotion by means of combining several musical features commonly used for music classification. With this representation, a computer algorithm should be able to determine the kind of emotion a human subject would be most likely to experience give a certain piece of music.

MATERIALS AND METHODS

Psychological emotion models: In psychology and philosophy, emotion is a subjective, conscious experience characterized primarily by psychophysiological expressions, biological reactions and mental states (Cherry, 2013). Many studies attempted to determine which are the basic emotions. There are some theories about music emotion classification. Two of the most popular approaches of music classification are categorical (Hevner, 1936) and dimensional (Thayer, 1989) models of

Corresponding Author: Weina Zhu, [1]School of Information Science, Yunnan University, China [2]Kunming Institute of Zoology, Chinese Academy of Science, China

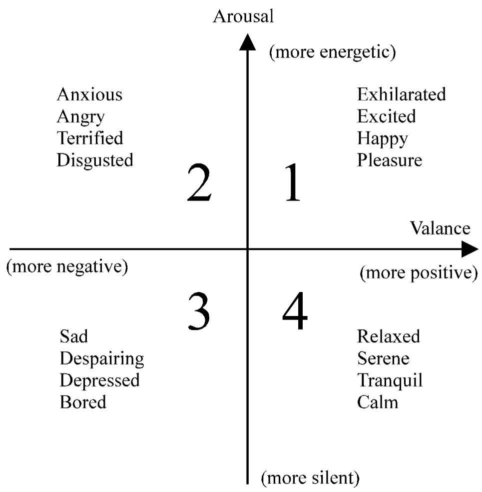

Fig. 1: Two-dimensional emotion representation in Thayer's model (Thayer, 1989)

emotion. The categorical approach describes emotions with a limited number of basic and universal categories such as happy, merry, sad, dreamy, tender, anger and fear, etc. The dimensional model forms an emotion plane which entails two factors: valence (negative to positive) and arousal (calm to exciting). Comparing to the categorical model, the emotion plane encompasses all affective terms arising from independent neurophysiological systems (Thayer, 1989). As shown in Fig. 1, there are four quadrants on the emotion plane and each quadrant corresponds to a set of emotion categories.

To use the dimensional model for classification, we asked subjects score arousal and valence values separately (from 1-10) for each of a set of music pieces by questionnaire. Then each music piece was represented as a point in the emotion plane which classified the music pieces into four categories.

FORMAL REPRESENTATION OF EMOTION

The relationship between emotion feature (E) and other classic music features (f_i) is defined as Eq. 1. n is the number of classic features we used to "construct" emotion.

$$E = \{f_1, f_2, \ldots f_i \ldots f_j \ldots f_n\} \tag{1}$$

In Eq. 1, all classic music features equally devote to the representation of emotion. However it has been proven that the contribution of each feature to emotion is different (Hevner, 1936; Kaminska and Woolf, 2000; Dalla Bella *et al*., 2001; Geringer and Madsen, 2003; Hailstone *et al*., 2009). Therefore, we improved the Eq. 1 as 2. w_i is the weight of each classic feature:

$$E = f(f_1w_1, f_2w_2, \ldots f_iw_i, \ldots f_jw_j, \ldots f_nw_n) \tag{2}$$

FEATURE EXTRACTION

Based on previous studies, pitch, key, melody, rhyme and timbre are most important features that influence the emotion content of music. Therefore, we picked them as f_i in Eq. 2. Because each of the features is distinct from each other, we used different methods to extract them from the music pieces. We used MIRtoolbox (Lartillot and Toiviainen, 2007) to extracted the features.

f_1: Pitch: Pitch is a perceptual property that allows the ordering of sounds on a frequency-related scale (Klapuri and Davy, 2006). Pitches can be perceived as "higher" and "lower" in the sense associated with musical melodies (Plack *et al*., 2005). Therefore, we used pitch interval (f_1) as the pitch related feature of music. In musical set theory, a pitch interval is the number of semitones that separates one pitch from another, upward or downward (Schuijer, 2008). Equation 3 notated pitch interval (f_1) of two continuous pitch p_{i+1} and p_i.

$$f_1 = p_{i+1}-p_i \tag{3}$$

f_2: Rhythm: Rhythm generally means a movement marked by the regulated succession of strong and weak elements, or of opposite or different conditions. In music, rhythm is the timing of events on a human scale of musical sounds and silences (Jirousek,1995). We used the ratio of speed as the feature of rhythm as in Eq. 4. Here l_i denoted the duration of individual tones:

$$f_2 = \frac{(l_{i+2} - l_{i+1}) - (l_{i+1} - l_i)}{(l_{i+1} - l_i) - (l_i - l_{i-1})} \tag{4}$$

f_3: Melody: A melody is a temporal succession of musical tones that the listener perceives as a single entity which is a combination of pitch and rhythm (Apel, 1944). We use the averaged pitch difference and averaged relative duration difference as the "melody" feature as in Eq. 3 (Chen *et al*., 2007). Here p_i is the pitch and l_i is the duration of individual tones:

$$f_3 = \left\{ \frac{\sum_{i=1}^{n-1} p_{i+1} - p_i}{n}, \frac{\sum_{i=1}^{n-1} \frac{l_{i+1} - l_i}{l_i}}{n} \right\} \tag{5}$$

f_4: Timbre: Timbre is the quality of a musical sound that distinguishes different types of musical instruments (e.g., string instruments, wind instruments and percussion instruments, etc.), even when they have the same pitch and loudness. The physical characteristics of sound that determine the perception of timbre include spectrum and envelope. We defined timbre by Eq. 7. Here m (t) is the waveform function of music associated with time (t). By doing a Fourier transformation of m (t), we got the power spectral density as the feature of timbre. (ω is radian frequency):

$$f_4 = \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} m(t) e^{-i\omega t} dt \right|^2 \tag{6}$$

f_5: Key: In music theory, the key of a music piece usually refers to the tonic note and chord which gives a subjective sense of arrival and rest. Major and minor are frequently referred to the key of a piece. Classical music is usually in a key. Here we use Eq. 4 to define the feature of key.n:

$$f_5 = \begin{cases} 1 & \text{when} \quad \text{mior} \\ 2 & \text{when} \quad \text{major} \end{cases} \tag{7}$$

By replacing the features (f_i) we extracted from above process, given a music piece including n notes, we updated Eq. 2 as:

$$E = f((f_{11}, f_{12} ... f_{1n})w_1, (f_{21}, f_{22} ... f_{2n})w_2, f_3w_3, f_4w_4, f_5w_5) \tag{8}$$

Here, we have two n-Dimensional descriptive vectors to denote pitch and rhythm features of a piece of music with n notes, respectively. In order to compare between different music pieces, we use mean and standard deviation value of f_1 (pitch) and f_2 (rhythm) as in Eq. 9:

$$E = f((f_{11}, f_{12})w_1, (f_{21}, f_{22})w_2, f_3w_3, f_4w_4, f_5w_5) \tag{9}$$

CLASSIFICATION

We used two binary classifications (there exist exactly two groups, exactly one of which a sample belongs to) in this article which are due to the nature of the emotion plane. One classification was used on valence (negative vs. positive) and another one was used on arousal (calm vs. exciting).

Because linear classification has a comparably low computational cost and good performance on linearly separable datasets, in this article, we used a linear classifier (Tabachnick and Fidell, 2000), similar to a discriminant analysis. For a given sample S, the score of the classification function C_j for one of the two possible groups j = {1|2} is found by multiplying the values in each dimension of the data space by their associated, learned coefficients c_{jn}:

$$C_j = c_{jo} + \sum_{n=1}^{N} c_{jn} S_n \tag{10}$$

With n = 1: N iterating the dimensions of the data space. The coefficients c_{jn} are learned from the means and the pooled covariance matrix of the data dimensions.

First, the selected classification algorithm is trained on the music pieces database which was labeled by human subjects on valence and arousal values separately. Second, a number of samples are used to test the learned classifier and each sample is assigned to the group corresponding to the result of the evaluation. In order to avoid "overlearning", a procedure called "cross-validation" is used: for a k-fold cross-validation, the entire available data is arranged in k equally sized subsets. Then, k-1 of these are used for training and the remaining 1 is used for testing. The particular subset used for testing is cycled through exactly k repetitions, so that every subset has been tested once (Drewes, 2006).

Classification accuracy is then reported as the mean of all of the classification accuracies on the k testing subsets. Result reliability increases with larger k, but so does the computational complexity due to the number of training/testing cycles. The demand for more reliable results and computational feasibility therefore need to be balanced out, limiting the amount of cross-validations used to values that are usually far below the theoretically possible maximum (which would be N, the number of samples available).

CONCLUSION AND DISCUSSION

By experiments on real music pieces, we extracted the emotion factor of all music pieces and classified them into different emotion categories. We found that the effectiveness of the pitch feature for musical emotion classification was impaired by using mean and standard deviation values. In other studies the pitch feature was used as a n-D feature vector, but in our work, in order to compare between music pieces with different numbers of notes (different "n" values), we had to unify the number of features. In future work, we will focus on resolving this conflict.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (61005087, 61263042), Science and Technology Project of Yunnan Province, China (2009CD018), Key Science and Technology Project of Department of Education, Yunnan Province, China (2010Z067).

REFERENCES

Apel, W., 1944. Harvard Dictionary of Music. Harvard University Press, Cambridge, ISBN-10: 0674011635, pp: 517-519.

Chen, X.Y., F. Gao and R.Z. Lu, 2007. A new indexing method for content-based music information retrieval and its application. Comput. Eng. Appl., 43: 233-235.

Cherry, K., 2013. Theories of emotion. Major Theories of Emotion. http://psychology.about.com/od/psychologytopics/a/theories-of-emotion.htm

Dalla Bella, S., I. Peretz, L. Rousseau and N. Gosselin, 2001. A developmental study of the affective value of tempo and mode in music. Cognition, 80: B1-10.

Drewes, J., 2006. Classification of natural scenes. Ph.D. Thesis, Giessen University, Giessen.

Geringer, J.M. and C.K. Madsen, 2003. Gradual tempo change and aesthetic responses of music majors. Int. J. Music Educ., 40: 3-15.

Ghias, A., J. Logan, D. Chamberlin and B.C. Smith, 1995. Query by humming-musical information retrieval in an audio database. Proceedings of the 3rd ACM International Conference on Multimedia, November 5-9, 1995, Francisco, California, pp: 231-236.

Hailstone, J.C., R. Omar, S.M. Henley, C. Frost, M.G. Kenward and J.D. Warren, 2009. It's not what you play, it's how you play it: Timbre affects perception of emotion in music. Q. J. Exp. Psychol. (Colchester), 62: 2141-2155.

Hevner, K., 1936. Experimental studies of the elements of expression in music. Am. J. Psychol., 48: 246-268.

Huron, D., 2000. Perceptual and cognitive applications in music information retrieval. Proceeding of the 1st International Symposium on Music Information Retrieval, October 23-25, 2000, Plymouth, Massachusetts, USA., pp: 11-12.

Jirousek, C., 1995. Rhythm. Art, Design and Visual Thinking. http://char.txa.cornell.edu/language/principl/rhythm/rhythm.htm

Kaminska, Z. and J. Woolf, 2000. Melodic line and emotion: Cooke's theory revisited. Psychol. Music, 28: 133-153.

Klapuri, A. and M. Davy, 2006. Signal Processing Methods for Music Transcription. Springer, USA., ISBN: 978-0-387-30667-4, Pages: 8.

Lartillot, O. and P. Toiviainen, 2007. A matlab toolbox for musical feature extraction from audio. Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07), September 10-15, 2007, Bordeaux, France, pp: 1-8.

Naoko, K., N. Yuichi, S. Tetsuo, Y. Masashi and K. Kazuhiko, 2000. A practical query-by-humming system for a large music database. Proceedings of the 8th ACM International Conference on Multimedia, October 30- November 3, 2000, Los Angeles, CA., USA., pp: 333-342.

Plack, C.J., A.J. Oxenham and R.R. Fay, 2005. Pitch: Neural Coding and Perception. Vol. 24, Springer, USA., ISBN-13: 978-0387234724, Pages: 364.

McNab, R.J. and L.A. Smith, 1996. Melody transcription for interactive applications. Working Paper, 96/32, Department of Computer Science, University of Waikato, Hamilton, New Zealand, pp: 20-32. http://researchcommons.waikato.ac.nz/handle/10289/1194.

McNab, R.J., L.A. Smith and I.H. Witten, 1996a. Signal processing for melody transcription. Proceedings of the 19th Australasian Computer Science Conference, January 31-February 2, 1996, Melbourne, Australia, pp: 301-307.

McNab, R.J., L.A. Smith and I.H. Witten, C.L. Henderson and S.J. Cunningham, 1996b. Towards the digital music library: Tune retrieval from acoustic input. Proceedings of the 1st ACM International Conference on Digital Libraries, March 20-23, 1996, ACM, Bethesda, Maryland, USA., pp: 11-18.

Schuijer, M., 2008. Analyzing Atonal Music: Pitch-Class Set Theory and Its Contexts. University of Rochester Press, Rochester, NY., ISBN: 978-1-58046-270-9, pp: 35.

Sonoda, T. and Y. Muraoka, 2000. A WWW-based melody retrieval system: An indexing method for a large database. Proceedings of the International Computer Music Conference, (CMC'00), Brlin, Germany, pp: 170-173.

Tabachnick, B.G. and L.S. Fidell, 2000. Using Multivariate Statistics. Allyn and Bacon, Boston, ISBN-13: 978-0205849574, pp: 1024.

Thayer, R.E., 1989. The Biopsychology of Mood and Arousal. Oxford University Press, USA., ISBN-13: 978-0195068276, Pages: 256.

Won, J.Y., J.H. Lee, K. Ku, J. Park and Y.S. Kim, 2005. A Content-Based Music Retrieval System Using Representative Melody Index from Music Databases. In: Computer Music Modeling and Retrieva, Wiil, U.K. (Ed.). Springer-Verlag, Berlin, Heidelberg, ISBN: 978-3-540-24458-5, pp: 280-294.

Ye, C. and Y. Jiang, 1988. Introduction of Musical Aesthetics. Beijing University Press, China, ISBN: 9787301003121, pp: 31-33.