Music Emotion Classification (MEC): Exploiting Vocal and Instrumental Sound Features

Mudiana Mokhsin Misron[1,*], Nurlaila Rosli[1],
Norehan Abdul Manaf[1], and Hamizan Abdul Halim[2]

[1] Faculty of Computer and Mathemathical Sciences
Universiti Teknologi MARA
Malaysia
[2] Faculty of Technology Management
Open Universiti Malaysia
Malaysia
{mudiana,norehan}@tmsk.uitm.edu.my,
laila8805@gmail.com,
hamizan.abdhalim@tm.com.my

Abstract. Music conveys and evokes feeling. Many studies that correlate music with emotion have been done as people nowadays often prefer to listen to a certain song that suits their moods or emotion .This project present works on classifying emotion in music by exploiting vocal and instrumental part of a song. The final system is able to use musical features extracted from vocal part and instrumental part of a song, such as spectral centroid, spectral rolloff and zero-cross as to classify whether selected Malay popular music contain "sad" or "happy" emotion. Fuzzy *k*-NN (FKNN) and artificial neural network (ANN) are used in this system as a machine classifier. The percentages of emotion classified in Malay popular songs are expected to be higher when both features are applied.

1 Introduction

Music has become more and more important in human lives and the need to improve the development of music acquisition and storage technology keep on rising. Music is a super-stimulus for the perception of musicality, where musicality is a perceived aspect of speech that provides information about the speaker's internal mental state [1]. It is believed that violation of or conformity to expectancy when listening to music is one of the main sources of musical emotion [2]. Thus, it is essential to conduct research as to analyze the similarities among music pieces based on which music can be organized in groups and recommended to user with suitable tastes. According to [3], music classification studies have so far been done with the main focused on classifying music according to genre and artist style. Recently, the

* Corresponding author.

T. Herawan et al. (eds.), *Recent Advances on Soft Computing and Data Mining SCDM 2014*, Advances in Intelligent Systems and Computing 287,
DOI: 10.1007/978-3-319-07692-8_51,

affective or to be specific the emotion aspect of music has become one of the important criterions in music classification.

Music emotion classification (MEC) is part of music data mining and artificial intelligence (AI) area of science. According to oxforddictionaries.com, music can be defined as vocal or instrumental sound and its common elements are pitch, rhythm, dynamics and timbre. Whereas, emotion is refer to as a strong feeling deriving from one's circumstances, mood or relationship with others. Primary emotion classes are happiness, sadness, anger, surprise, disgust and fear [4]. Emotion in certain music can be classified by employing two main processes namely, signal modelling and pattern matching. Based on work done in [5], signal modelling is referred to method of translating music audio signal into a set of musical features parameters. While, pattern matching is the process of parameter sets discovery from memory which strongly matches the parameter set obtained from the input music audio signal. All of this process automatically carried out using AI machine classifier such as supervised vector machine (SVM), artificial neural network (ANN), decision tree and etc. [6].

Until recently, most of MEC is done by looking at features such as audio, lyrics, social tags or combination of two or more features as stated above [7-9]. However, there were only few studies on MEC that exploits features from vocal part of the song [8]. It has been proved that, the timbre of the singing voice, such as aggressive, breathy, gravelly, high-pitched, or rapping is often directly related to our emotion perception and important for valence perception [10] thus it is suggested that vocal timbre should be incorporated to MEC. This research is proposed, to develop emotion classification system for Malay popular music from the year of 2000-2013. The final system should be able to use musical features extracted from vocal part and background music of a song as well as able to classify the type of emotion in music. The system will be employing two classification techniques namely, artificial neural network and fuzzy k-NN in order to classify category of emotion in selected Malay popular music. The overall system has implying data mining classification algorithm and techniques based on "Soft Computing and Data Mining" technology.

The discussion in this paper is divided into five sections, where the first part explains the overall idea of this research. Part two illustrates the literature review where, the previous and related works is clarified. Part three describes the data collection. Part four explains about music emotion classification system setup, fuzzy k-NN, ANN training and testing and classification results of the study. Part five discuses conclusions and proposed future works.

2 Literature Review

Generally, there are four main important things that need to be considered and understood in audio based music emotion classifications. Numbers of factors might obstruct the construction of a database and the issues such as which emotional model or how many emotion categories should be used in order to generate data collection must first be decided before one can proceed to the initial phase of MEC [8]. Fig. 1 below illustrates the typical audio based MEC as taken from [11].

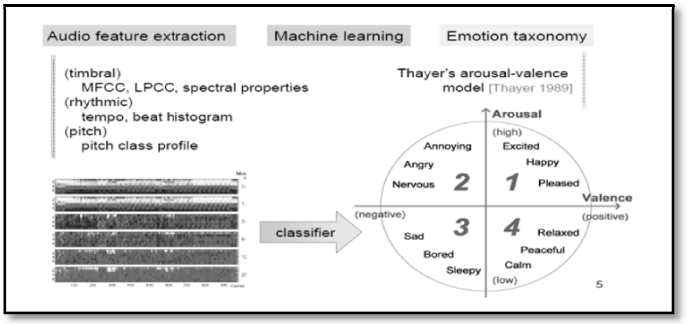

Fig. 1. Typical Audio Based Music Emotion Classification Taken from [11]

The growth in MEC have triggered the development of technology and system that related to emotion analysis and music signals, for example Moodtrack [12], Mood Cloud [13], and i.MTV [14]. These multimedia systems have applied MEC techniques which includes subjective test, musical features extraction, machine learning algorithm and etc. Normally, a subjective test is conducted to collect the ground truth data needed for training the computational model of emotion prediction. Subjective test can be done by numbers of annotation process, where selected annotators, manually listen to certain song and classify it based on group or classes.

Music listening is very subjective and multidimensional, especially in terms of emotions triggers. According to [15] and [16], different emotion insights of music are usually related with different patterns of acoustic cues. For example, excitement or feeling happy (arousal) is associated to tempo (fast/slow), pitch (high/low), loudness level (high/low), and timbre (bright/soft), where as sadness or valence is related to mode (major/minor) and harmony (consonant/dissonant) [17]. Generally, features of music such as timbre, rhythm, and harmony are extracted to signify the audio parameters of a music clips. Timbre is the characteristic of music stricture that can makes someone cry when a sad song being played to them. It is what makes a violin sound so beautifully sad and saxophone so blissfully happy. The timbre controls any emotion associated with the sound [18]. Several machine learning algorithms also applied to learn the relationship between music features and emotion labels.

The most used machine learning algorithms in MEC are artificial neural network (ANN). ANN has become one of the most significant mining techniques in various areas of science [19](Giudici,2003). The main reason for exploiting ANNs in those areas is because ANNs are very compatible as it can cater problem in various field and ANN is easy to manoeuvre as its operated just as same as human brain [20]. Another successful classification techniques in MEC besides ANN's, is by using the fuzzy k-NN (FKNN) classifier, as fuzzy logic is proved to be able to deal with uncertainty and imprecision in MEC [21]. Generally, two techniques has been used in this study, where the ANN classifier has been developed based on the overall MEC system, from training to testing, while, FKNN classifier technique has been developed based on work done in [22].

3 Data Collection

Due to the lack of ground truth data, most researchers compile their own databases [23]. Manual annotation is one of the most common ways to do this. For the purpose of this project, input music data are chose based on the result from subjective test that were carried out by categorizing Malay song into two main emotions, which is happy and sad. The rational of choosing only two emotions is because, happy and sad emotion is the basic emotion from psychological theories [24] and these two emotions cover a wide opposition of differentiation in 2D Rusell affect model representation with valence and arousal dimension [25]. "Happy" are positive valence and respectively high arousal, whereas "Sad" are in negative valance and respectively low arousal. So basically, both happy and sad quotient represents opposite value.

3.1 Subjective Test

Subjective Annotation test must been done in order to get all final 100 song with happy and sad emotion categorization. Previous studies have highlighted the important and common practice when doing this subjective test.

- Reducing the length of the music pieces [26][27].
- Providing synonyms to reduce the ambiguity of the affective terms [26].
- Using exemplar songs to better articulate what each emotion class means [28].
- Allowing the user to skip a song when none of the candidate emotion classes is appropriate to describe the affective content of the song [28].

For the purpose of this study, no restrict to exclusive categories will be compromised in order to undergone this subjective test. For each emotion, the problem is considered as binary classification. For example, one song can be categorized as either "happy or not happy" same goes to "sad or not sad". The dataset collection is made of 300 popular Malay song that is taken from Malay song charts from year 2000-2013 (Sources: Malay Radio Charts and "Anugerah Juara Lagu"). From this test, merely 50 songs that represent only "happy" emotion and another 50 songs represent only "sad" emotion will be allowed to be in the data collection for training purpose. To ensure the accuracy of the categorization process, 10 randomly selected annotators among teenagers with age range from 16-19 years old, were enquired to identify whether or not the data collection only contain "happy" and "sad" song.

3.2 Features Extraction

Music features extraction is the most crucial part in this study. It involves audio features extraction which has taking place as to determine the accuracy of data generation in the database. Generally, this project only focuses on timbre features which comprises of Spectral Rolloff, Zero-Cross, and Spectral Centroid.

Matlab programming is used to extract all of those selected features from every part of audio data (vocal part and instrumental part).

❖ *Spectral Rolloff*

Representation of the spectral shape of a sound and they are strongly correlated. It's defined as the frequency where 85% of the energy in the spectrum is below that frequency. If K is the bin that fulfills;

$$\sum_{n=0}^{k} x(n) = 0.85 \sum_{n=0}^{N-1} x(n) \quad (1)$$

Then the Spectral Rolloff frequency is f(K), where x(n)represents the magnitude of bin number n, and f(n) represents the center frequency of that bin.

❖ *Spectral Centroid*

Measure used in digital signal processing to exemplify a spectrum. It indicates where the "center of mass" of the spectrum is. Perceptually, it has a strong correlation with the impression of "brightness" of a sound. It is calculated as the weighted mean of the frequencies present in the signal, determined using a Fourier transform, with their magnitudes as the weights. Equation (2) is a formula to find the amount of spectral centroid in certain song.

$$Spectral \quad Centroid \quad = \frac{\sum_{k=1}^{N} kF[k]}{\sum_{k=1}^{N} F[k]} \quad (2)$$

❖ *Zero-Cross*

Zero-Cross is the number of times a sound signal crosses the x-axis, this accounts for noisiness in a signal and is calculated using the following equation (3), where sign is 1 for positive arguments and 0 for negative arguments. X[n] is the time domain signal for frame t.

$$Z_t = \frac{1}{2} \sum_{n=1}^{N} | sign(x[n]) - sign(s[n-]) \quad (3)$$

4 Music Emotion Classification System

In order to classify two types of emotion to be exact, sad and happy emotion in selected Malay popular music, music data first must be converted into a standard format specifically; 22,050 Hz sampling frequency, 16-bits precision, 30 second frames. Overall process of MEC system from developing and testing are using MATLAB R12 programming language.

4.1 Fuzzy *k*-NN (FKNN) Classifier

Fuzzy k-NN (FKNN) classifier has implied combination of fuzzy logic and k-NN classifier. FKNN is widely used in pattern recognition. In [22], fuzzy membership $\boldsymbol{\mu_{uc}}$ for an input sample $\mathbf{x_u}$ to each class $\mathbf{c}$ as a linear combination of the fuzzy vectors of k-nearest training samples. where $\boldsymbol{\mu_{ic}}$ is the fuzzy membership of a training sample $\mathbf{x_i}$ in class $\mathbf{c}$, $\mathbf{x_i}$ is one of the k-nearest samples, and $\mathbf{w_i}$ is the weight inversely proportional to the distance $\mathbf{d_{iu}}$ between $\mathbf{x_i}$ and $\mathbf{x_u}$:

$$\mu_{uc} = \frac{\sum_{i=1}^{k} w_i \mu_{ic}}{\sum_{i=1}^{k} w_i} \tag{4}$$

$$w_i = d_{iu}^{-2} \tag{5}$$

With Eq. (4), we get the C×1 fuzzy vector $\boldsymbol{\mu_u}$ indicating music emotion strength (C = 2) of the input sample:

$$\mu_u = \{\mu_{u1}, \ldots, \mu_{uc}, \ldots, \mu_{uC}\}^t \tag{6}$$

$$\sum_{c=1}^{C} \mu_{uc} = 1 \tag{7}$$

According to [22], the fuzzy vector of the training sample $\boldsymbol{\mu_i}$ is computed in fuzzy labeling section. Several methods have been developed in [21] and [29], where $\mathbf{v}$ is the voted class of $\mathbf{x_i}$, $\mathbf{n_c}$ is the number of samples that belong to class $\mathbf{c}$ in the K-nearest training samples of $\mathbf{x_i}$, and $\boldsymbol{\beta}$ is a bias parameter indicating how $\mathbf{v}$ takes part in the labeling process ($\boldsymbol{\beta} \in [0,1]$). Different $\boldsymbol{\beta}$ is used during cross validation process, ($\boldsymbol{\beta}$=0.0, 0.25, 0.50, 0.75, 1.0). When $\boldsymbol{\beta}$=**1**, this is the crisp labeling that assigns each training sample full membership in the voted class $\mathbf{v}$. When $\boldsymbol{\beta}$=**0**, the memberships are assigned according to the K-nearest neighbors. The equation can be generalized as:

$$\mu_{ic} = \begin{cases} \beta + (n_c / K) * (1 - \beta), & \text{if } c = v. \\ (n_c / K) * (1 - \beta), & \text{if } c \neq v. \end{cases} \tag{8}$$

4.2 Artificial Neural Network (ANN)

The concept of Artificial Neural Networks (ANN) is based on biological neural networks. Neural network approaches have shown to be promising in supporting fundamental theoretical and practical research in artificial intelligence [20].

4.2.1 Neural Network Training

The network architecture used in this research is the feed forward back propagation. Neural network toolbox in MATLAB was utilized for training the neural network. It includes several variations of the standard back propagation. A variable learning rate that is a combination of adaptive learning rate and momentum training is used to train music clips data. 100 vocal audio data (comprises with 50 "sad" and 50 "happy" songs) and another 100 instrumental audio data (also comprises with 50 "sad" and 50 "happy" songs) data were used to train the neural network. All training data were in the standardized audio format. Training data was obtained from various sources in the internet and Malaysia's radio station. All of this audio data are split into 30 second frames.

4.2.2 Neural Network Testing

Testing process in MEC take place after database comprises with musical features are generated. Music data says for example "*Ombak Rindu"* one of the Malay popular songs is entered to the system. Automatically system will extract musical features from that particular song before ANN classifier can classify category of emotion contained in the song.

During the classification process, ANN classifier will get the information from the database or (memorized value of musical features) from previous training process. ANN classifier then can classify emotion from the song by scheming the music features vector as to produced result that close to 1 (happy) or close to 0 (sad).

As shown in Fig. 2, songs with an output ranging from $0.5 \leq x \leq 0.9$ were considered as happy songs, while songs with output less than $0.5 \geq x \geq 0$ were considered as sad songs. These tests were further verified using neural networks.

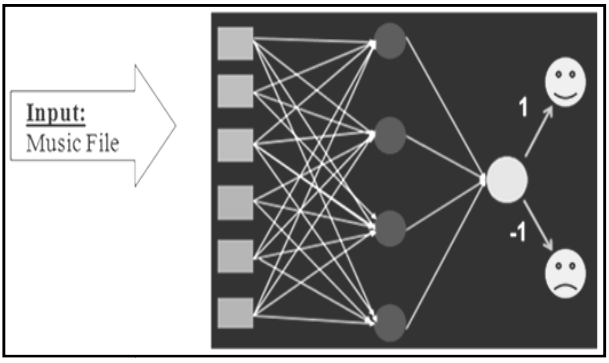

Fig. 2. ANN model for audio data testing

4.3 Testing Using Different Data

For system performance and classification accuracy evaluation, both classifier FKNN and ANN is tested with different data features. The testing process can be done using

three different algorithms. This is to see the differences in classification rate when using different data. The details of algorithm used are as follow:-

- FKNN/ANN + Only Vocal Features
- FKNN/ANN + Only Instrumental Sound Features
- FKNN/ANN + Vocal + Instrumental Sound Features

4.4 Experimental Result

4.4.1 Cross Validation Result for Fuzzy k-NN Classifier Using Different β

Table 1. FKNN Classifier Using Different β

β	*Happy→Happy*	*Sad→Sad*	*Average*
0.0	78%	43%	60.5%
0.25	81%	46%	63.5%
0.50	72%	46%	59%
0.75	84%	57%	70.5%
1.0	87%	51%	69%

4.4.2 Result Classification Accuracy

30 songs that were categorized as happy song and the other 30 songs categorized as sad song were used to test the algorithm. Summary of the results is shown in Table 2 and Table 3.

Table 2. Test Results

Description	*No. of Data*	*Using FKNN %*	*No. of Data*	*Using ANN%*
Happy song	30	100	30	100
Classified as Happy Song	15	50	26	86.6
Sad Song	30	100	30	100
Classified as sad song	16	53.3	24	80

The accuracy of the classification result can be measured by dividing number of correctly classified songs with the total number of songs. A comparison of the accuracy of using only vocal features and the combination of vocal and instrumental sound features is shown in Table 3. The tests were administered using the same set of test music. Results show that the proposed approach which is using both data is more competitive than using only vocal or instrumental features as training data.

Table 3. Classification Rate Using Different Training Data

Algorithm	*Using FKNN %Accuracy*	*Using ANN %Accuracy*
Only Vocal Features	51	72
Only Instrumental Sound Features	52	75
Vocal+ Instrumental Sound Features	53.3	83.3

Based on the results, the accuracy of the algorithm is higher (more than 80%) when using ANN classifier for both vocal and instrumental sound features. Whereas, the accuracy of the algorithm using FKNN shows quite positive results although only able to classify slightly half of the selected song with exact emotions. It is shown that, the highest accuracy is at 70.5% when using $\beta = 0.75$.

5 Conclusion and Future Works

The music classification algorithm developed is proven to be up to 80% accurate using ANN techniques, while, the percentages of emotion successfully classified using FKNN is approximately 50%. The manoeuvring of vocal and instrumental features with the assistance of ANN classifier can provide successful music emotion classification. Data from timbre extraction for both vocal and instrumental sound is used as training data to the neural network. Vocal and instrumental sound features were combined to improve testing and classification accuracy. ANN learns to recognize emotion in music based on timbre musical texture as exist in the database. The system is developed through learning rather than programming. However, ANN is still unpredictable. It may take some time to learn a sudden drastic change. As for the fuzzy k-NN classifier, generally FKNN classifier has successfully classified songs in regards to certain group of emotions though the results not as high as when using ANN.

5.1 Future Works

Overall, this project has been manipulating two basic emotions as to categorize emotion in selected music (happy and sad affects). Besides, this work only focus on extracting timbre vectors in the music data, in which previous studies have recommend that timbre can be used to strongly determined the emotion or behaviour in both vocal and instrumental sound data. As for machine classifier, this project has used one of the most well-known artificial intelligence machines learning to be precise, Artificial Neural Network (ANN) and also fuzzy classifier (FKNN). Both of these techniques had been proved to be able to generate positive result, as expected. However, for future study, it is suggested that another types of music excerpt such as pitch, energy,

harmony and etc; can be used to improved musical features database for training and testing process. With the positive result congregates from this project, it is extremely recommended if other types of emotion be considered as part of the classification category. This will hope to improve music emotion classification in the future.

Acknowledgments. This work is supported by a grant from the RMI, Mara University of Technology, Shah Alam.

References

1. Dorell, P.: What is Music?: Solving a Scientific Mystery, 318 p. NZ Publishing, Wellington (2005)
2. Imbrasaite, V.: Absolute Or Relative? A New Approach To Building Feature VecTors For Emotion Tracking In Music. In: Luck, G., Brabant, O. (eds.) Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, June 11-15 (2013)
3. Hu, Y., Chen, X., Yang, D.: Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. In: Proceedings of the International Conference on Music Information Retrieval (2009)
4. Picard, R.W., Vyzas And, E., Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Trans. Pattern Anal. Mach. Intell. 10, 1175–1191 (2001)
5. Gilkes, M., Kachare, P., Kothalikar, R., Pius, V., Pednekar, R.M.: MFCC-based Vocal Emotion Recognition Using ANN. In: International Conference on Electronics Engineering and Informatics (ICEEI 2012) IPCSIT, vol. 49 (2012)
6. Lartillot, O., Toiviainen, P.: A Matlab Toolbox for Music Information Retrieval. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (eds.) Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 261–268 (2008)
7. Hu, Y., Chen, X., Yang, D.: Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. In: Proceedings of the International Conference on Music Information Retrieval (2009)
8. Yang, Y.H., Chen, H.H.: Machine Recognition of Music Emotion: A Review. ACM Transactions on Intelligent Systems and Technology 3(3), Article 40 (2012)
9. Xu, M., Duan, L.-y., Cai, J., Chia, L.-T., Xu, C.S., Tian, Q.: HMM-based audio keyword generation. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3333, pp. 566–574. Springer, Heidelberg (2004)
10. Turnbull, D., Barrington, L., Torres, D.: Semantic annotation and retrieval of music and sound effects. IEEE Trans. Audio, Speech Lang. Process. 16(2), 467–476 (2008)
11. Yang, Y.-H., Lin, Y.-C., Cheng, H.-T., Liao, I.-B., Ho, Y.-C., Chen, H.H.: Toward multi-modal music emotion classification. In: Huang, Y.-M.R., Xu, C., Cheng, K.-S., Yang, J.-F.K., Swamy, M.N.S., Li, S., Ding, J.-W. (eds.) PCM 2008. LNCS, vol. 5353, pp. 70–79. Springer, Heidelberg (2008)
12. Vercoe, G.S.: Moodtrack: practical methods for assembling emotion-driven music. M.S. thesis, MIT,Cambridge, MA (2006)
13. Laurier, C., Herrera, P.: Mood cloud: A real-time musicmood visualization tool. In: Proceedings of the Computer Music Modeling and Retrieval (2008)

14. Zhang, S., Qingming, H., Qi, T., Shuqiang, J., Wen, G.: i. MTV: an integrated system for mtv affective analysis. In: Proceedings of the 16th ACM International Conference on Multimedia, pp. 985–986. ACM (2008)
15. Krumhansl, C.L.: Music: A link between cognition and emotion. Current Directions in Psychological Science 11(2), 45–50 (2002)
16. Juslin, P.N.: Cue utilization in communication of emotion in music performance: relating performance to perception. Journal of Experimental Psychology: Human Perception and Performance 26(6), 1797 (2000)
17. Gabrielsson, A., Erik, L.: The influence of musical structure on emotional expression (2001)
18. Lakatos, S.: A Common Perceptual Space for Harmonic and Percussive Timbres. Perception & Psychophysics 62(7), 1426–1439, PMID 11143454 (2000)
19. Giudici, P.: Applied Data Mining: Statistical Methods for Business and Industry. John Wiley & Sons, Inc. (2003)
20. Zurada, J.K.: Introduction to Artificial Neural Systems, 2nd edn. West Publishing Company (2006)
21. Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy k-nearest neighbor algorithm. IEEE Transactions on Systems, Man and Cybernetics (4), 580–585 (1985)
22. Yang, Y.H., Liu, C.C., Chen, H.H.: Music emotion classification: a fuzzy approach. In: Proceedings of the 14th Annual ACM International Conference on Multimedia. ACM (2006)
23. Yang, D., Lee, W.S.: Disambiguating Music Emotion Using Software Agents. In: ISMIR, vol. 4, pp. 218–223 (2004)
24. Juslin, P.N., Sloboda, J.A.: Music and emotion: Theory and research. Oxford University Press (2001)
25. Russell, J.A.: A circumplex model of affect. J. Personal. Social Psychol. 39(6), 1161–1178 (1980)
26. Skowronek, J., Mckinney, M.F., Van De Par, S.: A demonstrator for automatic music mood estimation. In: Proceedings of the International Conference on Music Information Retrieval (2007)
27. Yang, Y.-H., Lin, Y.-C., Cheng, H.-T., Liao, I.-B., Ho, Y.-C., Chen, H.H.: Toward multi-modal music emotion classification. In: Huang, Y.-M.R., Xu, C., Cheng, K.-S., Yang, J.-F.K., Swamy, M.N.S., Li, S., Ding, J.-W. (eds.) PCM 2008. LNCS, vol. 5353, pp. 70–79. Springer, Heidelberg (2008)
28. Hu, Y., Chen, X., Yang, D.: Lyric-based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method. In: ISMIR, pp. 123–128 (2008)
29. Han, J.H., Kim, Y.K.: A Fuzzy K-NN Algorithm Using Weights from the Variance of Membership Values. In: CVPR (1999)