2018 2nd International Conference on Informatics and Computational Sciences (ICICoS)

Feature Selection for Music Emotion Recognition

Emilia Widiyanti
Department of Informatics/Computer Science
Faculty of Science and Mathematics, Universitas Diponegoro
Semarang, Indonesia
emilia.widiyanti@gmail.com

Sukmawati Nur Endah
Department of Informatics/Computer Science
Faculty of Science and Mathematics, Universitas Diponegoro
Semarang, Indonesia
sukma_ne@undip.ac.id

***Abstract*— Feature selection is step in preprocessing that can be used to reduce data dimension and eliminate the irrelevan data. There are several algorithms in feature selection. This study will compare several feature selection algorithms, namely Sequential Forward Selection, Sequential Backward Selection, and Relief F to find features that are very influential in musical emotional recognition. The method in music emotion recognition uses Support Vector Machine with the RBF kernel. The experimental results show that based on the recognition results with the highest accuracy, the most influential features are the zero crossing rate, music mode, harmonics, pitch and energy obtained through the Sequential Backward Selection algorithm. The selection of features in this study can increase accuracy up to 8%.**

Keywords—feature selection, music emotion recognition, Sequential Forward Selection, Sequential Backward Selection, ReliefF

I. Introduction

Music Emotion Recognition (MER) is a topic focusing on emotional content of music. Researches on this topic are various, and affect other topics such as Music Information Retrieval, Music Understanding, and many more. Researchers strive to develop their researches on this topic by doing many experiments with the music content, feature extraction, machine learning, and feature selection.

Feature selection is one of many step in pre-processing that can be used to reduce data dimension, eliminate the irrelevan data and increase the performance of accuracy [1]. By using feature selection to analize the big data content, more effective work could be done where the algorithm should not process the unusable features on the data.

There are many algorithms in feature selection. The usage of selection feature is based on types and dimensions of the data [2]. Recognition process in MER can be based on regression problem or classification problem. Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) [3] are feature selection algorithms that can be used both in regression or classification problem. There are also other feature selection algorithms that focused only in single type of problem such as ReliefF that focused on multiclassification problem and RReliefF that focused on regression problem [4]. This paper compares several selection feature algorithms to find features that are influential in MER.

In this paper, 13 music features (such as zero crossing rate, spectral centroid, spectral entropy, spectral flux, spectral rolloff, energy, key clarity, musical mode, chroma deviation, harmonics, pitch, tempo, and tempo regularity) are extracted from dataset. Then using selection feature, best features of music are selected. The selection feature algorithms used in this paper are SBS, SFS, and ReliefF that are focused in classification problem. This paper use Support Vector Regression as classifier and train the model based on features selected by feature selection algorithm. Performance evaluation is done by compute the accuracy of model produced .

This paper consists of four sections. Section 1 describes a brief introduction about this study. Section 2 explains the research methodology. Section 3 presents experiment and result. In the final section, we conclude this study.

II. Methodology

First, the dataset is composed by doing data collection process. The complete dataset will be processed through the data preprocessing to be uniformed as needed. After preprocessed, then the music features of the dataset could be extracted.

Extracted features of dataset will be processed as training and testing data after going through the selection feature algorithm. Training process or model building will produce model that will be used in order to recognize the testing data. Then, the model evaluation of the recognition process could be done by compute the accuracy of the models. Figure 1 describes the process flow of methodology for this study.

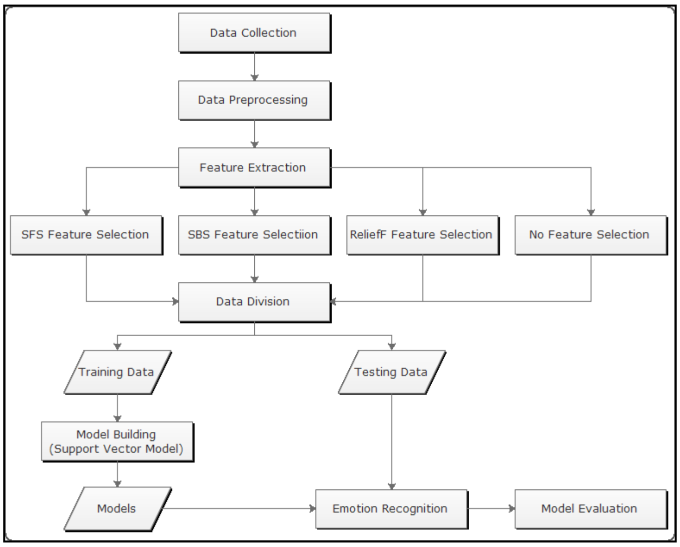

Fig. 1. The Process Flow of Research Methodology

Detail of process can be explained as follows.

A. Data Collection

The music database used for the research is instrumental songs from movie soundtrack. The collection is done by collecting the songs from music websites such as www.neosounds(.)com that provide the song categories which could help in identifying the emotion. Five basic emotion categories: angry, nervous, happy, neutral, and sad, as has been used in similar research before [5]. So, the dataset is composed of 20 songs for each emotion, in total of 100 songs for the entire dataset for five emotions considered.

978-1-5386-7440-6/18/$31.00 ©2018 IEEE
Authorized licensed use limited to: Universidade de Coimbra. Downloaded on January 19,2023 at 21:28:08 UTC from IEEE Xplore. Restrictions apply.

B. Data Preprocessing

The composed dataset needs to be processed before feature extraction. There are several processes that should be done. First, cut the data in 30 second length per data in the dataset. Cutting process done manually using Audacity opensource software. Then, uniform the dataset as WAV file format with mono channel. Last, do the resampling for the dataset in 22050 Hz. The file format and channel uniform, and the resampling process are done using Freemake Audio Converter opensource software.

C. Features Extraction

This paper extracts various music features as main feature data. These **features** are features that are widely used in research about music emotion recognition.

1) Zero Crossing Rate

Zero crossing rate is the rate of sign-chances of the signal during the duration of particular frame [6]. This paper computes the standar deviation of zero crossing rate as the representation of zero crossing rate feature[7].

2) Spectrum Features

Spectral features are features computed from the STFT of an audio signal. **Spectral centroid** means the centroid of the magnitude spectrum of STFT,

$$\text{spectral centroid} = \frac{\sum_{n=1}^{N} nA_t^n}{\sum_{n=1}^{N} A_t^n} \quad (1)$$

where A_t^n, is the magitude of the spectrum at the *t*-th frame and the n-th frequency bin and N is the total number of bins [7].

Spectral rolloff is defined as the frequency K_t below which a certain fraction of the total energy is contained [7].

$$\sum_{n=1}^{K_t} A_t^n = 0.85 * \sum_{n=1}^{N} A_t^n \quad (2)$$

Spectral Flux is defined as the square of the difference between the normalized magnitudes of successive frames [7]. And the last, **Spectral entropy** is define as entropy of the normalized spectral energies for a set of sub-frames. Each spectral feature is represented using computation of the standard deviation[6].

3) Energy

To determine energy feature of signal, we could compute standard deviation of the **average energy** of the overall signal or wave sequence. It is define as:

$$\text{Average energy}(x) = \frac{1}{n}\sum_{t=0}^{n} x(t)^2, \quad (3)$$

$$\sigma\big(\text{Average energy}(x)\big) = \sqrt{\frac{1}{n}\sum_{t=0}^{n}(AE(x) - x(t))^2} \quad (4)$$

where variable x represents the signal, variable t represents time in samples and variable n represents the length or numbers of x in samples [8].

4) Key Clarity, Musical Mode, Chroma Deviation

Key clarity and musical mode of the music or signal is computed by applying key-finding algorithm by krumhansl [9]. Chroma feature is used here as musical representation of the signal. Chroma feature has 12 elements and represents the summed chromagram. In this paper, standard deviation of chroma feature is used to represent the **chroma deviation.**

Tonality of the signal computes by extracting chroma feature of the signal then combines it with key characterization by the following equation [8]:

$$\text{Tonality} = \text{vC.KPM} \quad (5)$$

$$\text{Key} = \max_{\text{keyIndex}}(\text{Tonality}(\text{Idx})) \quad (6)$$

where, vC is chroma feature and KPM is key profile matrix by Krumhansl. By computing tonality, **key clarity** is defined as the index of key in tonality with the highest strength and the **musical mode** is define as the difference between the best major key and the best minor key in tonality [7].

5) Harmonics

Harmonics feature is estimate by computing the harmonics distribution yeilds (HS). A method to compute harmonic distribution yields [8] define as:

$$HS(f) = \sum_{k=1}^{M} \min(\|X(f)\|, \|X(kf)\|) \quad (7)$$

where variable M represents the maximum number of harmonics, then variable f represents the fundamental frequency, and X represent the STFT of the signal. In this paper, harmonics feature is defined by computing the standard deviation of HS computed beforehand.

6) Pitch

Pitch is also defined as perceived fundamental frequency can be extracted using many method. In this paper, pitch is extracted using autocorrelation method. **Pitch** is estimated by computing the autocorrelation of the signal, then estimate the peak of the autocorrelation computed using parabolic function.

7) Rythmn Features

Tempo is defined as the beats per minute (BPM) which is used to represent the global rhythmic feature of music [8]. Tempo and beat feature is extracted by using tempo analysis and beat tracking that proposed by Ellis et al [10] (Ellis and Poliner 2007). The overall tempo is extracted to represent **tempo** feature of the music. Then, **tempo regularity** is defined by extracting the beats of the data then compute the standard deviation of the beats intervals [8].

D. Feature Selection

This paper compares several feature selection algorithms including Sequential Forward Selection, Sequential Backward Selection, and ReliefF. In addition, this study will compare the result when selection feature are not applied.

1) Sequential Forward Selection (SFS)

Sequential forward selection is feature selection algorithm that utilizes a particular learning method to evaluate possible feature subsets based on the performance result of that particular learner [11]. SFS is the simplest greedy search algorithm and performs best when the optima subset has a small number of features [12]. This algorithm sequentially adds feature based on the results of the objective function.

Algorithm [12]:

1. Compose an empty set $Y_\theta = \{\phi\}$ to contain the selected feature
2. Selecting the best feature $X^B = \arg\max[J(Y_k+X)]; x \notin Y_k$

Authorized licensed use limited to: b-on: Universidade de Coimbra. Downloaded on January 19,2023 at 21:28:08 UTC from IEEE Xplore. Restrictions apply.

3. Updating the selected features container $Y_{k+1} = Y_k + X^B$; k=k+1
4. Do 2

where Y is set of selected features, X is the features, X^B is the best feature selected in iteration, k is indeks iteration, and $[J(Y_k+X)]$ is objective function of SFS.

2) Sequential Backward Selection (SBS)

Different with SFS that start with empty set, sequential backward selection do the opposite by starting with full set of features. This algorithm sequentially remove the feature that result smallest decrease in the value of the objective function [12].

Algorithm [12]:

1. Compose a set $Y_\theta = X$ to contain all features
2. Removing the worst feature selected $X^W = \text{argmax}[J(Y_k\text{-}X)];x \notin Y_k$
3. Updating $Y_{k+1} = Y_k - X^W$; k=k+1
4. Do 2

Where $J(Y_k\text{-}X)$ is the objective function of SBS, and X^W is worst feature selected.

3) ReliefF

The ReliefF is the extension of basic feature selection Relief. ReliefF method works on the multiple-class problem and this algorithm aims at estimating the quality of features according to how well their values separate the instance according their distance in the problem space [11]. ReliefF function describe on [4] is applied in this paper to determine the rank of the data features.

E. Data Division

The data divided into training data and testing data. Training data is used to train the model and the testing data is used for emotion recognition data. The data is divided by applying k-fold cross validation method.

The data is divided in k part of data. Each part of data is used as testing data sequentially as the other is used as training data. In this study, the value of k is defined as 10.

F. Model Building

Model building is a process to build a recognition model based on given training data by applying a machine learning method. SVM is one of many supervised learning models that can be used for classification or regression problem.

There are several parameters that are used in determining model using SVM. First parameter is known as kernel function. Kernel is a function to transform the problem given into a higher dimension. By using kernel function, the data that hard to be separated in the actual dimension is transformed into a high dimensional data where it could be more separable. The other parameter is Cost (C) or regularization parameter. This parameter controls the margin of the hyperplane.

This paper uses Support Vector Machine (SVM) as machine learning model and apply Radial Basis Function RBF) as kernel function. The model building process uses the training data as learning input and produces a model based on the training data and parameters.

G. Emotion Recognition

In this process, the models produced by model building or training process is used to predict the emotion of data. The data used in this process is the training data.

H. Model Evaluation

After doing the recognition (prediction), the result could be evaluated to see the performance of the models. In this paper, accuracy is used to be the evaluation parameter of models.

III. Experiments and Result

A. Experiments

In this section several experiment scenarios is done to find the best features or the influential features for MER using each feature selection. In order to find the best model, different value of cost and feature selected are assigned. The model building or training process is done as list in the assigned value of cost and features. The value of the cost that used are 1 (C1), 100 (C2), 10000 (C3), and 1000000 (C4) and the RBF kernel parameter used in this paper is assigned as 0.01.

In this paper, k-fold cross validation is applied and the value of k is assigned as 10 to avoid over-fitting problem.

First Scenario

Sequential Forward Selection is applied in this scenario as the feature selection algorithm. The selection feature algorithm computation is done for each possible number of features. For each number of features, based on the features selected, build models for values of cost assigned before. Last, compute the accuracy of each model. The evaluation result is shown in Figure 2.

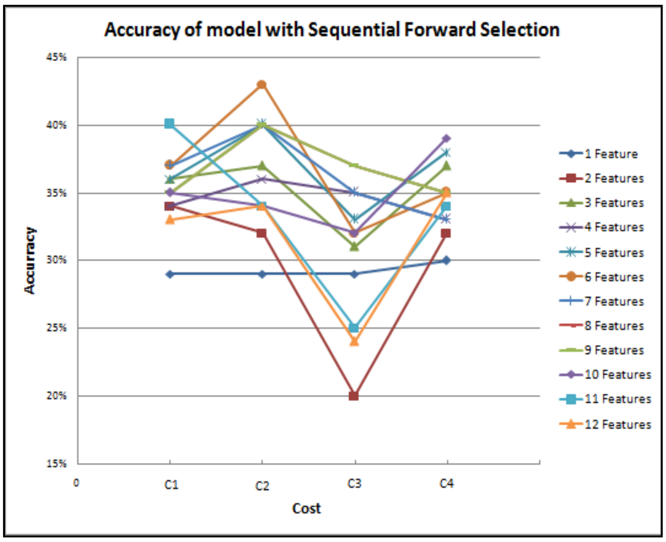

Fig. 2. Accuracy of model with Sequential Forward Selection

Authorized licensed use limited to: b-on: Universidade de Coimbra. Downloaded on January 19,2023 at 21:28:08 UTC from IEEE Xplore. Restrictions apply.

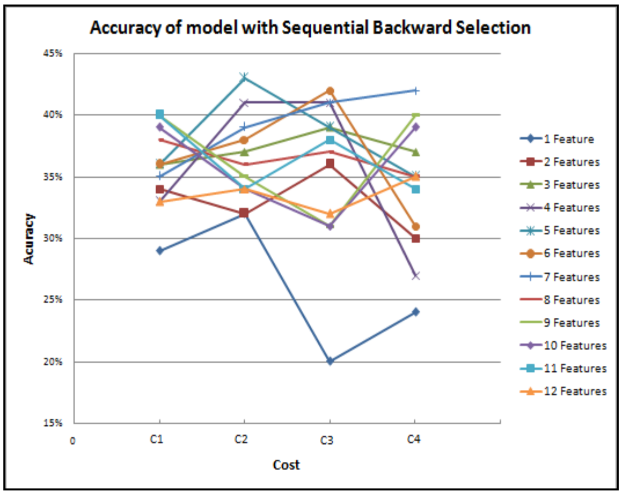

Fig. 3. Accuracy of model with Sequential Backward Selection

Second Scenario

Sequential backward selection is applied in this scenario as the feature selection algorithm. The steps done in this scenario are the same as describe in the first scenario. The result of computation shown in figure 3.

Third Scenario

In this scenario, the ReliefF feature selection algorithm are applied. Slightly different with the two scenario described before, ReliefF algorithm result is a rank of features. So, in this scenario, for each model the features used are choosen based on the rank. The result of computation in this scenario is shown in figure 4.

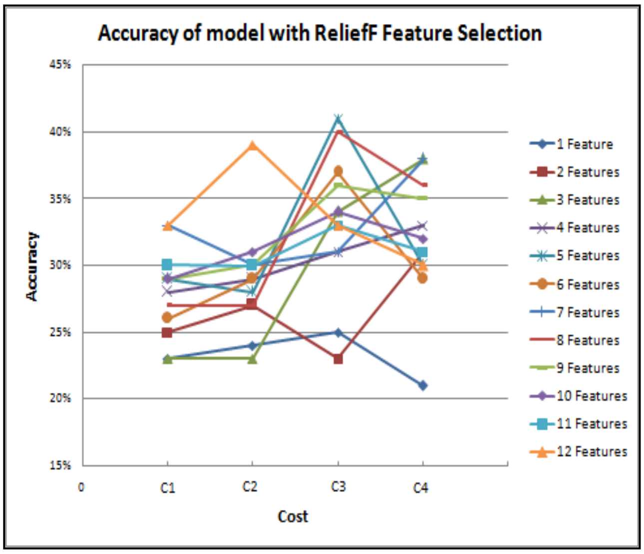

Fig. 4. Accuracy of model with ReliefF Feature Selection

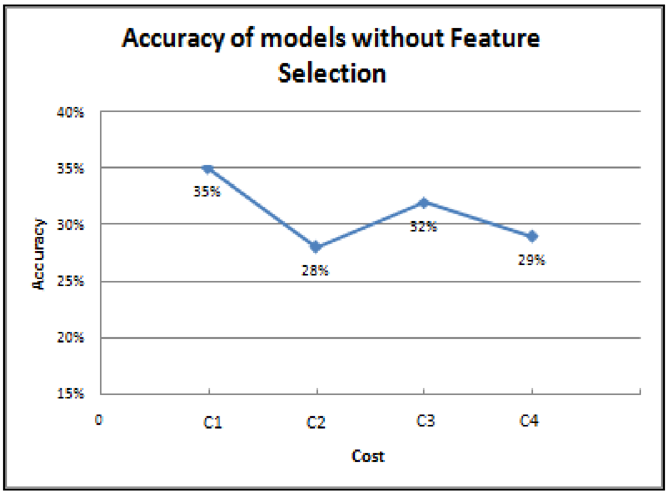

Fig. 5. Accuracy of models without Feature Selection

TABLE I. COMPARISON BETWEEN FEATURE SELECTION ALGORITHMS

Feature Selection	Best Accuracy	Number of features	Feature Selected
SFS	43%	6	zero crossing rate, key clarity, musical mode, harmonics, pitch, energy
SBS	43%	5	zero crossing rate, musical mode, harmonics, pitch, energy
ReliefF	41%	5	spectral entropy, key clarity, musical mode, harmonics, tempo
No Feature Selection	35%	13	zero crossing rate, spectral centroid, spectral entropy, spectral flux, spectral rolloff, chroma deviation, key clarity, musical mode, harmonics, pitch, energy, tempo, beats

Fourth Scenario

In this scenario, any feature selection are not applied and the computation is done using all features extracted. The result is shown in figure 5.

Fifth Scenario

This scenario compares the result of the first to fourth scenarios. It takes the best accuracy that result in each scenario. The result can be seen in Table I:

B. Result Analysis

Features that influence music emotional recognition are the features with the highest recognition results. Based on Table I, it can be seen that there are five features that can produce the highest accuracy of all the features tested. These features are zero crossing rate, musical mode, harmonics, pitch and energy.

The features selected show a close connection with music emotion recognized. One of the features selected is energy feature. Energy feature is a feature that shows how loud the energy on the music. The value of this feature will influence the music emotion recognized. Recognition of music with high value of energy feature will result in the emotions that has high arousal such as happy and angry, while the low

Authorized licensed use limited to: b-on: Universidade de Coimbra. Downloaded on January 19,2023 at 21:28:08 UTC from IEEE Xplore. Restrictions apply.

value of energy feature will result in the emotion that has low arousal such as sad. This connection also could be seen in other features selected that is zero crossing rate, musical mode, harmonics, pitch and energy. These features are very influential in recognizing the emotion of music.

The results of this feature selection are obtained from the use of the SBS algorithm. Although SFS has the same accuracy as SBS, the number of features produced by SBS is less than SFS. Therefore in this study using SBS is more recommended because it can reduce the cost of computation time.

By comparing the evaluation results of emotion recognition without selection feature and using selection features, it shows that the usage of the selection features can increase the accuracy by 8%.

IV. CONCLUSION

Based on the experimental results, it can be concluded that the most influential feature on musical emotional recognition is zero crossing rate, musical mode, harmonics, pitch and energy. This feature is obtained by using the SBS algorithm. The selection of features in this study can increase accuracy by 8%.

The value of the accuracy of emotional recognition using SVM is still relatively low. Thus, in future work it is suggested to analyzr other methods that can increase accuracy.

REFERENCES

[1] L. Yu, and H. Liu, “Feature Selectionfor High Dimensional Data: A Fast Correlation-Based Filter Solution”, Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003

[2] A. Jovic, K. Brkic, and N. Bogunovic, “A review of feature selection methods with applications”, 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 25-29 May, 2015

[3] A. Huq, J.P Bello2, and R. Rowe2, Automated Music Emotion Recognition: A Systematic Evaluation, Journal of New Music Research, 2010, Vol. 39, No. 3, pp. 227–244

[4] M. R. Sikonja and I. Kononenko, “Theoretical and empirical analysis of relieff and rrelieff”, 2003, Machine Learning, Vol. 53, No. 1-2 pp. 23-69

[5] N. J. Nalini and S. Palanivel, “Emotion recognition in music signal using aann and svm”, 2013, International Journal of Computer Applications, Vol. 77, Issue 2

[6] T.Giannakopoulos and G. Pavan, “PyAudioAnalysis: an open-source python library for audio signal analysis”, 2015, PloS one.

[7] Y. S. Yang and H. H. Chen, “Music emotion recognition”, USA, CRC Press, 2011

[8] B.-j. Han, S. Rho, R. B. Dannenberg, and E. Hwang, “Smers: music emotion recognition using support vector regression”, 2009, ISMIR, pp. 651-656

[9] C. L. Krumhansl, “Cognitive foundations of musical pitch”, 1990, Oxford University Press

[10] D. P. Ellis and G. E. Poliner, “Identifying 'cover songs' with chroma feature and dynamic programming beat tracking”, 2007, IEEE Conf. on ICASSP, Vol. 4, pp. 1429-1432

[11] L. Cehovin and Z. Bosnic, “Empirical evaluation of feature selection methods in classification”, 2010, Intelligent Data Analysis, Vol. 14, Issue 3, pp. 265-281

[12] L. Ladha, T. Deepa, “Feature Selection Methods and Algorithms”, 2011, International Journal of Advanced Trends in Computer Science and Engineering, 3, pp. 1787-1797

.

Authorized licensed use limited to: b-on: Universidade de Coimbra. Downloaded on January 19,2023 at 21:28:08 UTC from IEEE Xplore. Restrictions apply.