January - February 2020
ISSN: 0193 - 4120 Page No. 85 - 93

An OMPGW Method for Feature Extraction in Automatic Music Mood Classification Using PSO-SVM

Shashi Shekhar, Vishal Goyal
Department of Computer Engineering & Applications
Department of Electronics & Communication Engineering
GLA, University, Mathura, India-281406
shashi.shekhar@gla.ac.in , vishal.goyal@gla.ac.in

Article Info
Volume 82
Page Number: 85 - 93
Publication Issue:
January-February 2020

Abstract

In music related applications, Music mood are very useful. Music signal's inherent emotional expression are represented music mood. Analysis of Music signal based on emotions are proposed in this paper. Wigner distribution function, Gabor functions and orthogonal matching pursuit forms base of this method. This method is termed as OMPGW and it has three schemes. They are high, middle and low level scheme. Music signal's adaptive time-frequency decomposition is provided by combining Gabor function with orthogonal matching pursuit in middle low-level schemes.

High temporal and spatial resolutions are given by proposed algorithm when compared to other algorithms. It also provides better representation of structure of music signal. From low-level schemes, results time-frequency energy distribution is obtained by applying Wigner distribution function in middle level schemes. Music mood classification procedure and audio feature modelling are described by high level schemes. Features are modelled using support vector machines with Particle Swarm Optimization classifier. Based on emotion model, features are extracted by proposed method.

Particle swarm optimization is proposed to search optimum parameters. It enhances the support vector machine's ability in generalization and learning. Four datasets are used for conducting experimentation and proposed method produced a better results. Various mood clusters are formed by classifying music clips in music mood classification with mean accuracy of 69.53%.

Article History
Article Received: *14 March 2019*
Revised: *27 May 2019*
Accepted: *16 October 2019*
Publication: *01 January 2020*

Keywords: *time-frequency analysis, orthogonal matching pursuit, music mood classification, feature extraction, Affective computing, parameters selection, Particle Swarm Optimization, Wigner distribution function, Support Vector Machine*

1. Introduction

In retrieval of music information field, automatic music emotion classification is getting increasing attention. The user activities in this field are not only extremely diversified, but also consistently growing. The diversity comes from the fact that emotions classification set up certain relationships between music and its effect in human emotional state like happy, anger, sad, tender, etc. In addition, the growth is inevitable to increase the accessibility to music databases. As the quantity of music satisfy continues to disprove, the searching time is unexpectedly increasing. The solution of widespread music collection under different emotions could lead to a decrease in the information retrieval search time on the online system.

People used to listen music during leisure period of time. Atpresent there is more and more music on personalcomputers, in music libraries, and on the Internet. Emotions can be expressed by Music . The music that people listen to is governed by what mood they are in. The music characteristics like timbre, pitch, melody and rhythm plays an important role in human physiological and psychological functions, thus altering their mood.

People used to listen some light music after work time for relaxation and they used to listen music with fast tempo and strong beat, when they are in gymnasium. Music makes easy way of communication between peoples. They also used to keep memories, emotions and to share them. Due to the advancement in internet, huge amount of data is stored in personal computers. So, in order to mange music, play list integration, music browsing and classification techniques need to be developed.

In object listening, there are various time concordance. Emotion based music retrieval and classification gives better results than classification based on genre, tempo, album and artist. There are more than one music mood in few music genres.

Various mood of music is generated by distinct musical features. Musical features at low level are used for getting high accuracy. Based on this, changes in musical mood can be detected and segments are formed by dividing music clips. Those clips with similar features are grouped to form clusters.

Using various models and features, few research methods are developed for classification of genre and detection of tempo and beat. Strong but short experience corresponds to emotion and less and long experience corresponds to mood. So, in this paper, mood is used. In order to ensure the word usage with reference, words like emotional expression, emotion and affect are used.

Automatic mood classification is done using proposed OMPGW method. Wigner distribution function [11], Gabor functions [10] and Orthogonal Matching Pursuit (OMP) [9] are used in this work. Music signal's time varying behaviour is handled by combining Gabor functions with OMP algorithm. Transient and rhythmic components of music signal are represented by this. Music signal's distribution of time-frequency energy is obtained by applying Wigner distribution function on OMP.

Audio feature modelling is presented depends on processing methods. Procedure of music mood classification is followed by this. Features are extracted in a better way and classification performance of music mood can be enhanced. Linear expansion of atoms are formed by decomposing music signal's in the proposed method. From time-frequency dictionary atoms are selected by using OMP algorithm. Distribution of time-frequency energy of music signal is generated by adding selected atoms Wigner distributions.

Audio features are extracted by applying distribution of time frequency energy and by decomposing coefficients. Based on this features, form feature set, which is termed as ATF features. Classifier learning is obtained by using ATF features. Excellent performance is exhibited by a proposed method on four different datasets as shown by experimentation. In classification music mood, around 69.53% of mean accuracy is achieved.

Following summarizes the major contributions of this work.

1) For classification of music mood, novel method OMPGW is proposed. Music signal's adaptive time-varying description is provided by this method. It provides high temporal and spatial resolution. Three signal processing algorithms are combined in this method and extraction of features are done by this.
2) Ten audio features are extracted based OMPGW which is termed as ATF features. Features includes, coefficient histogram, frequency cepstrum coefficient, sub-band power, spectral contrast, spectral flatness measure, spectral contrast, spectral

January - February 2020
ISSN: 0193 - 4120 Page No. 85 - 93

bandwidth, spectral flux, spectral rolloff, spectral centroid. Better performance are produced by proposed ATF features.

3) For classification of music mood, five datasets are used to evaluate proposed method. They are MediaEval 2015, MTV dataset, MIREX-T dataset, Soundtracks dataset. From MIREX-like mood dataset, adapt MIREX-T dataset. They are classified into four classes of mood.

Paper is organized as follows: Related work in music classification of audio features and in emotion models are reviewed in section 2. Section 3 describes proposed OMPGW music mood classification method. Section 4 discuss about experimentation and performance analysis of proposed method. Section 5 presents conclusion with future enhancement.

2. Related work

Chen et al. [1] used deep Gaussian process to implement a system to detect music emotion. There are two major stages. They are classification and extraction of feature. For emotion recognition, use deep Gaussian process in classification stage. In Gaussian process model, structural learning is provided by deep Gaussian process. In statistical machine learning, with deep architecture, learning methods, efficiency and potential are proved.

Bargaje et al. [2] optimized feature set for detecting emotions in audio samples. Genetic algorithms are used to reduce time of computation. Eight emotions are classified by modified 3D (Arousal-Valence-Loudness) plane from 2D (Arousal-Valence) type plane. Maik et al. [3] predicted emotion in two-dimensional Valence-Arousal space continuously using recurrent neural networks and stacked convolutional network. CNN layer with two trained RNN are used for arousal and valence.

Ascalon et al. [4] used lyrics to recognize music mood. Single or two or three sources are used in most of studies. Relation among mood and lyrics are investigated by this work. Using various parameters and threshold keyGraphkeyword generation algorithm and TF-IDF are applied.

‘Happiest.’, ‘Happy’, ‘Neutral,’, ‘Sad,’‘Saddest,’ are the most common classification of emotion. Various classifier’s are used for recognizing emotions of music according to different acoustic features. Classifier’s includes, support vector machines [7,8], support vector regression [6] and K-nearest neighbors [5].

In small and medium sized enterprises, intelligent IoT applications are developed by identifying new features of music. Also applied a combination of support vector machine and regression methods.

Proposed method is validated using emotion as induced by survey. While listening music, two class of emotions may occur. They are induced and perceived emotions.

Source conveys the perceived emotion. Evolution of state of art method is monitored by Music Information Retrieval Evaluation eXchange (MIREX). Efficiency of audio-based deep learning is shown by Lidy et al. [10]. , Jeon et al. [11] used bimodal convolutional recurrent network for presenting first multimodal deep learning method. It includes the representation of binary mood. Classical methods are used for making comparison. Mid level and late fusions are compared.

Induced emotion is subjective. Listeners may not be made happy by happy songs. Perceived emotion of songs are recognized by using various studies. It is objective and it is context invariant. Human feelings are used to derive emotion. Based on induced emotion, proposed method is validated.

3. Proposed Methodology

The framework of the proposed music mood classification system using the OMPGW method is shown in Fig. 1. This system contains three main parts: (1) proposed method is used to analyse signal , after pre-processing music clip. Music signal is decomposed adaptively by using OMP algorithm to form a Gabor function's linear expansion. Selected Gabor functions are added with Wigner distributions in order to compute music signal's time-frequency energy distribution.

From time-frequency energy distribution and decomposition coefficients, audio features are extracted. Using SVM-PSO classifier, music mood classification is applied with proposed ATF.

Figure 1 shows the framework of proposed classification system for music mood. There are three major stages.

(2) From every music clip which is given as input, compute OMPGW based ATF features. On every feature value normalization is applied.

(3) For PSO-SVMs, Thayers model of mood is concatenated with ATF features. Extract same features in testing and pre-trained PSO-SVMs is used to classify music clip's mood.

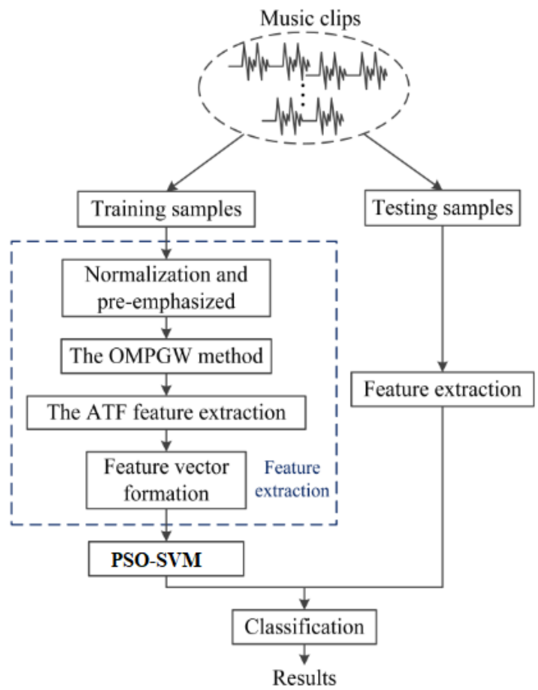

Fig. 1. Framework of the proposed music mood classification system

The OMPGW Method:

OMPGW method is used to represent music clips before the extraction of content-based audio feature. It is a signal processing technique. Signal's inherent structure can be obtained by this. Music signal's adaptive time-frequency estimate is generated by proposed method which is presented in this section. Wigner distribution function, Gabor functions and OMP algorithm forms base for this proposed method. Music signal is decomposed adaptively by using OMP algorithm to form a Gabor function's linear expansion. Selected Gabor functions are added with Wigner distributions in order to compute music signal's time-frequency energy distribution.

From time-frequency energy distribution and decomposition coefficients, audio features are extracted. Using SVM-PSO classifier, music mood classification is applied with proposed ATF. Music signal's Non- stationary nature is matched by proposed method and high spatial and temporal resolutions are given by proposed method when compared to other algorithms. Accuracy of mood classification of music can be enhanced by features of ATF.

When compared with vocal sounds of human, musical sound in wide frequency band are complex. Music signal has short transient and stationary part which will fade slowly. So, for analysis of music signals, time-frequency methods are used commonly. In classification system of music mood, best extraction of feature is done by proposed OMPGW method. It is an adaptive time-frequency analysis method. Compared to existing time-frequency analysis like MSTM, GTFB, CQT, Wavelet Transform, STFT, signal's multi-scale adaptive time-frequency decomposition is given by proposed OMPGW technique with Gabor function's redundant dictionary.

With arbitrary precision, compute signal's time and frequency components. Based on input signal they are selected. Structures of signal are matched in a better way by proposed method. Low time-frequency bandwidth product is contained by Gabor functions. It enhances the temporal and spectral resolution. For feature extraction, after OMPGW

January - February 2020
ISSN: 0193 - 4120 Page No. 85 - 93

application, outputs are saved. It includes, music signal's distribution of time-frequency energy and atoms with respective coefficients of decomposition. Following steps are in OMPGW method,

Algorithm. 1:

1) Initialization. Initialize residual error $R_0 f(t) = f(t)$for a music signal $f(t)$and selected set of atoms is $D_c = ø$. Assume iteration counter $k = 1$.

2) Atom selection. Compute inner product of $\{\langle R_k f(t), g_{\gamma k}(t)i\rangle;\ g_{\gamma k}(t) \in D_a\}$ and atom $g_{ck}(t)$with large value of inner product is computed. Reorder the selected set of atoms $D_{ck} = D_{c(k-1)} \cup \{g_{ck}(t)\}$, and column corresponds to selected atom $g_{ck}(t)$ and in k position, where selected atom is columns.

3) Orthogonalization. $P_k = D_{ck}(D_{ck}^{*} D_{ck})^{-1} D_{ck}^{*}$defines operator of orthogonal projection with columns span of D_{ck}. To residual error $R_k f(t) = (I - P_k) R_k f(t)$, orthogonal projection operator P_k is applied, where, identity matrix is represented as I.

4) Update. Update the model.$f(t) = \sum_{n=0}^{k-1} \langle R_n f(t), g_{cn}(t)\rangle$; $g_{cn}(t)$and $Rf_k(t) = f(t) - f_k(t)$.

5) Convergence. If stopping condition is satisfied, algorithm is converged. For example, small constant is greater than residual error. Else, increment the iteration counter k by one and repeat step1 -5.

6) Outputs. Obtain distribution of time-frequency energy $f(t)$ by adding Wigner distribution with every selected atom. Save respective coefficient $a_k = f(t), f_k(t)$ of every iteration.

Feature Extraction: In this section, OMPGW method is introduced to extract ATF features. From every input music signal, extract ATF features at first. From this features, construct a novel feature vector. With conventional processing methods, mood classification is done using these vectors. Processing methods includes, modulation spectral analysis, cepstral analysis. In following sub-section, feature extraction process is addressed in a detail manner.

ATF Feature Extraction: Down sample the original music signals in proposed system of mood classification in order to form uniform format. 16 bit resolution and 16KHz sampling rate is used. Before extraction of feature, segmentation of signals are not required by OMPGW. It is because of its adaptive time-frequency nature. Formation of frames by dividing every music signal is not required and it avoids ringing effect.

Spectral Centroid: Decomposition coefficients mass centredefines it. Similar to Fourier spectrum, with brightness impression, it has a robust connection. Spectral Rolloff: It is defined as decomposition coefficients below which magnitude distribution is concentrated.

Spectral Bandwidth: In a continuous set of frequencies, difference between upper and lower frequencies defines bandwidth.

Spectral Contrast: Difference between spectral valley and spectral peak defines it. Compute ATF spectral contrast.

Spectral Flatness Measure: It is also termed as tonality coefficient. Audio spectrum is characterized by this.

Spectral Crest Factor: It is used to differentiate tone and noise like sounds with respect to its spectral shapes. Flatness is inversely proportional to this.

Sub-band Power: For distribution of time-frequency energy, three sub-band power sections are computed.

Coefficient Histogram: variable's probability distribution can be estimated efficiently by histogram method.

Frequency Cepstrum Coefficient: Signal's inverse Discrete Fourier Transform (DFT) of log magnitude of DFT defines cepstrum. Feature vectors are formed by extracted ATF features. Music clips are represented by this. From OMPGW results, 9+L+3×M features derived, which is illustrated in figure 4.

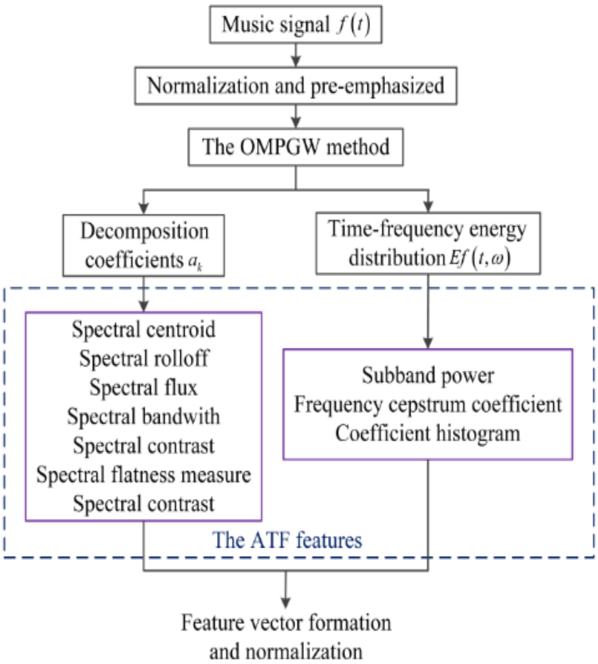

Fig. 2. Basic process of the ATF features extraction

1.1.Music Mood Classification

In classification of music mood system, classifier is applied after the extraction of features which are OMPGW based. Music clips are classified into various emotional classes like anxious, exuberance, depression and contentment by music mood classification. Proposed features are modelled as mood clusters using Support Vector Machines with PSO (SVM-PSO). With high accuracy, in classification of audio, SVM are used successfully.

SVM parameters optimization algorithm based on PSO:

In specified range, every parameter combination's accuracy estimate equals the estimate of parameter combination (C, γ) performance. Every parameter combination's accuracy is estimated using k cross validation. For optimization problem, best parameters are decided. Fitness value of every particle equals accuracy of k cross validation. Following shows the process of k cross validation. K folds are formed by separating the training data. For validation one fold is used and for training K-1 folds are used. Accuracy of k cross validation defines average accuracy value of validation set prediction. Every particle is a two dimensional (C, γ) one in optimization algorithm. Figure 3 shows this algorithm and it is described in as follows,

Algorithm: PSO based SVM parameters optimization

Input: Music feature data set

Output: optimal C and γ

Begin: 1) Initialization

- Value of C_1, C_2, w, particle number are set. Dimension range is specified. Particle's maximum speed also specified.
- Two dimension PSO population is established and for every particle, speed and location are initialized.
- Based on current location, every particle's personal extremes are initialized. For every particle, fitness value is computed and personal extreme of particle with best value of fitness is set as global extreme.

2) Until reaching stable solution or maximum count iteration, circulation is performed.

- Based on expressions (1) and (2), every particle's speed and location are updated.
- For every particle, fitness value is computed. If personal extreme's fitness value is less than new location's fitness value, personal extreme location is replaced by new location.
- If population's global extreme is worst then best particle, then global extreme is replaced by best particle.

3) Optimum value of C and γ are returned. End

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 85 - 93

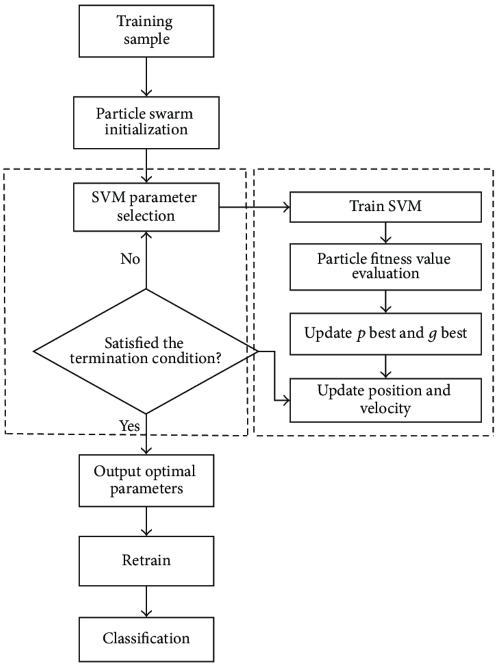

Fig.3.Flowchart of PSO-SVM classifier

4. Experiments and Results Discussion

Five mood annotated datasets are used for experimentation to evaluate proposed method's performance. Used datasets are described at first and presented the methods used to construct the experiments. With respect to accuracy of classification, proposed method is demonstrated. There are six classes of discrete mood in first dataset. They are tender, surprise, angry, fearful, sad and happy and this dataset is called as Soundtracks dataset. There 30 music clip records with 18 to 30 second duration. Randomly music clips are played. This dataset is publically available and most of the researchers uses this.

MIREX-like mood dataset is second dataset and it is re-annotated as Thayers mood of model. There are four mood classes in this dataset, which includes, anxious, exuberance, depression and contentment. It has 903 music clips with 30 seconds duration. All music clips are normalized to remove the effects of conditions in recording with unit variance and zero mean.

MTV database is a third database. It has 195 music clips with 14.2 h duration. Music in MTV database is selected from top ten of 20 years (1981C2000) of MTV Europe Most Wanted and it covers wide range of popular music genres. Four male and one female subjects are included in this with age from 22-23. With four classes, all songs are classified into Thayers mood model.

All votes of subjects are averaged to compute ground truth. There are two classes namely, valence and arousal. MediaEval 2015 database is a fourth database. It has 431 music lips with 45 second duration of every clip. Figure 2 shows the dimensional model. In music, emotions are described by employing this model on two orthogonal axes. They are arousal and valence. Annotators creates ground truth.

1.2.ROC Comparison Results

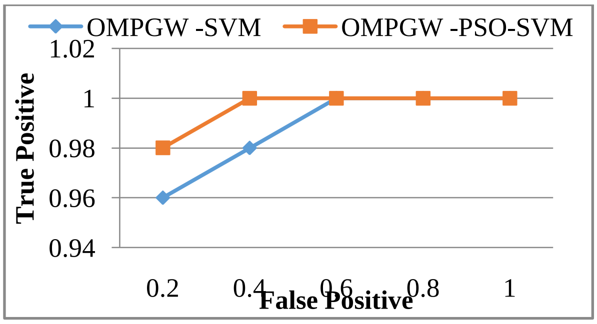

Fig.3.ROC curve comparison Results

Fig. 3 presents the ROC curves of proposed OMPGW -PSO-SVM and existing OMPGW -SVM. The existing method have achieved high performance in ROC curve, while OMPGW -PSO-SVM is a little higher. The high performance is as expected for the classification task here is similar to a basic classification problem in the unmodified scenario. It indicates that the classification is much easier because the data to be tested changes little from the template data. From the results it is concluded that the proposed OMPGW-PSO-SVM model is good for the classification of music clips.

January - February 2020
ISSN: 0193 - 4120 Page No. 85 - 93

1.3.Running Time Comparison Results

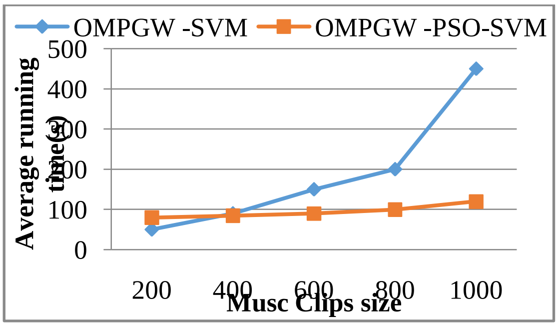

Fig.4.Running Time comparison Results

The running time of OMPGW -PSO-SVM and existing OMPGW -SVM are shown in Fig. 4. The time increases linearly as the size of images increases. It can also be seen from Fig. 4 that the running time of proposed method is much less than existing OMPGW-SVM. It can be concluded from the results that the proposed method can deal with large amounts of music clips data efficiently.

1.4.Detection Accuracy

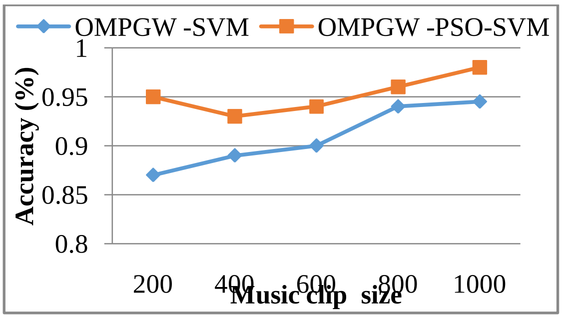

Fig.5.Detection Accuracy comparison Results

More specifically, Fig. 5 shows the average detection accuracy of proposed OMPGW-PSO-SVM and OMPGW-SVM, denoting the rate of correctly classified music mood cases. As denoted in Fig. 5, OMPGW-PSO-SVM achieves higher detection accuracy on various image sets with different sizes. The accuracy of OMPGW-PSO-SVM changes little with the increase of the size of image, while the accuracy of OMPGW-SVM fluctuate widely in some cases. This difference may be due to that the documents of same category in the test set and training set are quite different. OMPGW-PSO-SVM catches more feature information of images which could use the combined PSO and SVM.

1.5.Recall Comparison Results

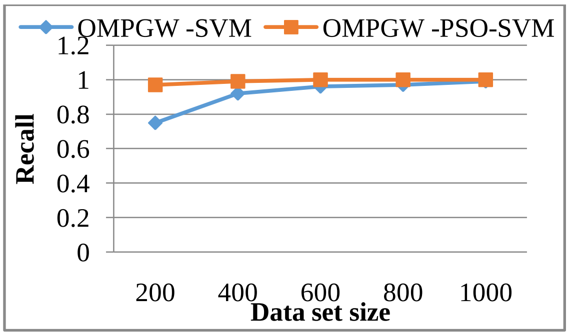

Fig.6.Recall comparison Results

The recall rates of OMPGW-PSO-SVM and OMPGW-SVM are provided in Fig. 6. As denoted in Fig. 6, OMPGW-PSO-SVM achieves a higher recall on various test image sets with different sizes. The recall of OMPGW-PSO-SVM rapidly increases little with the increase of the size of test set, while the recall of OMPGW -SVM lesser recall rate in some cases. This difference may be due to that the documents of same category in the test set and training set are quite different in proposed method uses PSO which has high convergence rate and thus improve the recall rate efficiently than the existing method.

5. Conclusion and Future work

Music mood classification is done by proposed OMPGW method in this research. It uses Wigner distribution function, Gabor functions and orthogonal matching pursuit. It has three schemes in music mood classification. Music signal is decomposed adaptively by using OMP algorithm to form a Gabor function's linear expansion. Selected Gabor functions are added with Wigner distributions in order to compute music signal's time-frequency energy distribution.

From time-frequency energy distribution and decomposition coefficients, audio features are extracted. Using SVM-PSO classifier, music mood classification is applied with proposed ATF. Music signal's Non- stationary nature is matched by

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 85 - 93

proposed method and high spatial and temporal resolutions are given by proposed method when compared to other algorithms. Accuracy of mood classification of music can be enhanced by features of ATF. Five mood annotated dataset is used for experimentation and it shows the efficiency of proposed method.

References

1. Chen, S.H.; Lee, Y.S.; Hsieh, W.C.; Wang, J.C. Music Emotion Recognition Using Deep Gaussian Process. In Proceedings of the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Hong Kong, China, 16–19 December 2015; pp. 495–498.
2. Bargaje, M. Emotion recognition and emotion based classification of audio using genetic algorithm-an optimized approach. In Proceedings of the 2015 International Conference on Industrial Instrumentation and Control, Pune, India, 28–30 May 2015; pp. 562–567.
3. Malik, M.; Adavanne, S.; Drossos, K.; Virtanen, T.; Ticha, D.; Jarina, R. Stacked convolutional and recurrent neural networks for music emotion recognition. In Proceedings of the 14th Sound and Music Computing Conference, Espoo, Finland, 5–8 July 2017; pp. 208–213.
4. Ascalon, E.I.V.; Cabredo, R. Lyric-based music mood recognition. In Proceedings of the DLSU Research Congress, Manila, Philippines, 2–4 March 2015; pp. 1–8.
5. Dewi, K.C.; Harjoko, A. Kid's Song Classification Based on Mood Parameters Using K-Nearest Neighbor Classification Method and Self Organizing Map. In Proceedings of the 2010 International Conference on Distributed Frameworks for Multimedia Applications, Yogyakarta, Indonesia, 2–3 August 2010; pp. 1–5.
6. Han, B.J.; Rho, S.M.; Dannenberg, R.B.; Hwang, E.J. SMRES: Music emotion recognition using support vector regression. In Proceedings of the 10th International Society for Music Information Retrieval Conference, Kobe, Japan, 26–30 October 2009; pp. 651–656.
7. Chang, C.Y.; Lo, C.Y.; Wang, C.J.; Chung, P.C. A music recommendation system with consideration of personal emotion. In Proceedings of the 2010 International Computer Symposium, Tainan, Taiwan, 16–18 December 2010; pp. 18–23.
8. Lin, C.; Liu, M.; Hsiung, W.; Jhang, J. Music emotion recognition based on two-level support vector classification. In Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, Jeju, South Korea, 10–13 July 2016; pp. 375–389.
9. Kawakami, A.; Furukawa, K.; Katahira, K.; Okanoya, K. Sad music induces pleasant emotion. Front. in Psychol. 2013, 4, 1–15.
10. Thomas Lidy and Alexander Schindler. Parallel convolutional neural networks for music genre and mood classification. MIREX, 2016.
11. Jeon, Byungsoo, Chanju Kim, Adrian Kim, Dongwon Kim, Jangyeon Park, and JungWoo Ha. "Music Emotion Recognition via End-to-End Multimodal Neural Networks." In *RecSys Posters*. 2017.