IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

A Historical View of the Progress in Music Mood Recognition

Swati Goel[1], Parichay Agrawal[2*], Sahil Singh[3], Prashant Sharma[4]

[1,2,3,4]Dept. of Computer Science and Engineering, ABES Institute of Technology, Ghaziabad, India

[*]Corresponding Author: parichayagrawal1998@gmail.com, Tel.: +91-7827787696

DOI: https://doi.org/10.26438/ijcse/v7i3.3945 | Available online at: www.ijcseonline.org

Accepted: 20/Mar/2019, Published: 31/Mar/2019

***Abstract*— This paper aims at assessing the state as well as the progress made in classifying emotions in the music. Music is known as “language of emotions”, hence its logical to consider it as a medium for determining the emotions as well as categorize the music based on the emotions they bring forth [1]. Different segments of a particular music may express different emotions and since emotions are interpreted by humans there may arise some conflicts to come to a well-defined answer. The ability to deduce the emotions exhibited by music is of great significance. For example, the ability to deduce emotions can help understanding the patients suffering from Alexithymia, online music vendors like Spotify, iTunes etc. can provide customized playlists based on moods. The task of emotion determination comes under the task of Music Information Retrieval henceforth referred to as MIR. The paper explores the methods of emotion retrieval that includes methods that use textual information (lyrics, tags etc.), content-based approaches and systems combining multiple methods [2].

***Keywords*—** Acoustic features, Music Emotion Recognition, MIREX, Social Tagging, MFCC, Centroid, Flux, Rolloff, Chroma, Gaussian Mixture Model, Support Vector Machine, VA Model, PAD Values

I. INTRODUCTION

Music and humans go back a long way to the past. According to the type of music, different music has various different kind of effects on humans. It is easier to quantify or subdivide songs on the basis of easily retrievable information like artist/ genre but MIR has received greater importance. Music on its own expresses emotions. These emotions could be interpreted differently hence it’s highly subjective and a difficult task [2]. Researchers have also shown that sound in organized form can resonate with one’s nerve tissues [3]. This relationship between music and the kind of emotions felt by humans has been studied by various researchers for a long time now. Barthet et al. [4] were the first to give detailed information about MER task. Wieczorkowska and others categorized this as a classification problem of multi-label type [5].

Now, with vast and easy to access music libraries over the internet, music is almost everywhere and the content is ever-increasing and hence the conventional methods used for managing these songs, i.e. the metadata (artist name, label name, song title etc.) approach would soon be insufficient. The use of mood conveyed by such songs would be a better approach which as of now is mostly done manually. The task of determining the mood of the music is interdisciplinary in nature which involves initial study of emotions, psychology as well as theory of music, signal processing and training of models to determine the mood.

In order to come up with standard methods for MIR systems and algorithms, Music Information Research Evaluation eXchange (MIREX) came to existence which later on accepted emotion recognition as a subtask of MIR in 2007 [6,7]. Under this annual meeting, the best of the systems demonstrate improvements each year. Though the scope of the features is limited to acoustic features but mood solely doesn’t depend on that. Many-a-times, lyrics also convey a particular type of emotion or a social issue which is reviewed here.

This paper aims to provide:

1. A consolidated view of the progress that has been made in the field of Music Information Retrieval (MIR).
2. Classify the different approaches used by the individuals and teams and compare their achievements.

The following content of the paper is organized as follows. Section 1 contains introduction of this paper, Section 2 describes the types of ‘Emotion Models’ that are used for mood recognition, Section 3 consists of methods that are used for emotion recognition, Section 4 discuss the indirect ways in which annotation of the music can be done by humans and Section 5 discusses the indirect ways for the same, i.e., through web scraping and other methods. Section 6 discusses the models and the need for the introduction of systems that annotate the music based on its content and shows why human annotation is tedious and redundant, Section 7 describes how the efficiency in mood recognition process can be improved by using a combination of models of different types (for example combining audio and tags. At last Section 8 concludes the review and highlights what is expected of MIR systems in the future.

II. EMOTION MODELS

Various psychological studies are conducted before a concrete structure is proposed to classify emotions. The emotion model can either be categorical or parametric in nature both of which are discussed as follows.

II.I **Categorical Model**

This approach involves identifying and sorting some emotion tags according to the relevance to some audio clip or music. In this area, an early study by Henver used 66 different adjectives (of moods) which were arranged into 8 groups [8]. Following the path Zentner *at al.* used 801 emotional terms and condensed it into a set of 146 terms that were unique for music mood identification [9].

For mood recognition, MIREX have given 5 categories/clusters where the songs can be categorized into any one of the same. These categories were derived by clustering matrix of labels. These labels (of popular songs) were obtained from All Music Guide [7].

Table 1: Mood adjectives according to MIREX mood classification task [7].

Clusters	**Mood Adjectives**
Cluster 1	Passionate, rousing, boisterous, rowdy, confident
Cluster 2	Rollicking, cheerful, fun, sweet, amiable/good natured
Cluster 3	Literate, poignant, wistful, bittersweet, autumnal, brooding
Cluster 4	Humorous, silly, campy, quirky, whimsical, witty, wry
Cluster 5	Aggressive, fiery, tense/anxious, intense, volatile, visceral

II.II **Spatial/Dimensional Model**

The other kind of model says that emotions can be laid down/ defined by a continuous plane of various labels. Major work in this field was done by Russell and Thayer who established a low dimensional model onto which these mood descriptors can be organized. The two dimensional model is the most prominent one which is a Valence-Arousal space where emotions are mapped along the arousal axis and valence axis. The arousal values are represented by y-axis and valence by x-axis. Arousal measures the intensity whereas Valence measures pleasantness of the music.

Studies have been made to expand the model to develop multi dimension models for more detailed analysis of the emotions. An attempt was made to expand the model into a three-dimensional model but has been a subject of disagreement [10]. As proposed by Thayer, there are possibilities of other dimensions as well like kinetics, tension, dominance that are not elements of valence [11].

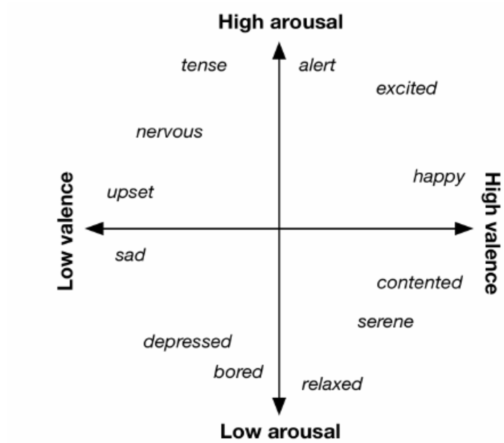

FIGURE 1: Thayer's 2D Model Diagram

III. EMOTION RECOGNITION METHODS

In order to evaluate the emotion of the song we attempt to annotate the music under consideration with a set of emotions. Thus, the emotion recognition can be thought of as a classification or regression problem. The song can either be or entire length or a section of the same. It could be of fixed length (20-30 seconds) or even as small as a 1-second clip.

The mood could be represented either of fixed type i.e., a single vector in multi-dimensional space or a time-series of vectors where each dimension of a vector represents an emotion. It is preferable to represent emotion as a time-series of vectors so as to determine the emotions at different durations of the song.

Emotion recognition can be done using various methods. The simplest one being the use of human listeners to check for the correctness of an emotion for a particular piece. Now this could be achieved either directly (survey, social tags) or indirectly (social tags, lyrics, web documents).

A more complex approach is to analyse the audio using signal processing and then using machine learning algorithms to automatically annotate the pieces with most relevant emotions. Lastly, multiple methodologies can be combined together to obtain a more refined result. For example, using audio and lyrics, audio and tags may be used in conjunction to determine the emotion of the audio in question.

IV. DIRECT HUMAN ANNOTATION

The techniques that comes under direct human annotation are surveys, social tags as well as tag games.

IV.I Surveys

Surveys are one of the most widely adopted and followed technique for annotating music with emotions. AMG has devoted their time and resources to obtain the emotional tags and aren't made open for public use. Hence the need was to come up with an open database for the same. Trohidis et el. made a dataset of 593 songs which were annotated by 3 expert listeners and each music was annotated by 6 emotions [12].

IV.II Social Tagging

Social tagging is another option which involves listeners in masses and a social platform where they can listen to the audio. Last.fm made significant contribution in this. It allowed users to contribute social tags to the songs they were listening to. By 2007, its users built up a social tag vocabulary of about 960,000 tags which were used to annotate thousands of songs [13]. Last.fm made its data available to the MIR community through its public APIs which was not so in the case of AMG. Below is the image of what social tags look like for a track in last.fm.

IV.III Annotation Games

The above methods are rather traditional and are time consuming, monotonous and expensive since it may involve hiring of expert listeners. A more refined approach is the use of interactive online games to collect labels for difficult problems. Many such games were made for the collection of data such as TagATune [14] and MajorMiner [15] where the focus have been the collection of data for short clips. MoodSwings is another game that used the Arousal-Valence model [16] in which a player partners with another player to place the cursor within the V-A space at the same time competing with others and the scores are calculated on the basis of the overlap between players' cursors.

V. INDIRECT HUMAN ANNOTATION

This section includes lyrics as well as web documents which are indirect sources of textual data from which emotions related to the music can be inferred.

V.I Web Documents

Many web pages containing artist or song reviews, biographies etc. are considered as rich sources for obtaining information regarding the emotions. Crawling a web page or blogs are techniques employed by many MIR systems that collect textual data from the internet [17]. Search engine queries are also used to obtain the information. These documents that are related to the songs are converted into a single document vector. The space thus obtained can be used to calculate music similarity. Knees et al. proposed a better technique than vector space called relevance scoring [18].

V.II Lyrics

The use of lyrics is comparatively more difficult than using tag based direct or indirect approaches. This is due to the fact that extracting the right set of words and labelling them to different classes of emotions is a rather difficult task. In order to use lyrics to identify the underlying emotions, the first task is to select those words from lyrics which have significant meaning to convey emotions. The second task is to use the database obtained to recognize different emotions i.e., to come up with systems to recognize emotions from lyrics. Lyrics are also used in combination with other features to obtain more accurate results.

V.II.I Word Selection

The major problem faced in using lyrics as a base to predict emotions is to find out the words that convey some emotions. The behavioural responses are caused by external stimuli. The behavioural responses are in turn due to emotional responses and can be described using Pleasure, Arousal and Dominance (PAD) values [19]. Following this study, a large set of words labelled with their PAD values called ANEW (Affective Norms for English Words) was given by Bradley [20]. ANEW was then used to develop ANCW (Affective Norms for Chinese Words) by Hu and others assuming that the translated words carry similar emotions as the English ones. For effective identification of features in lyrics BOW (Bag of Words) approach is used which takes the frequency of the words instead of the order in which those words appear. Obviously the naïve approach by Chen and others utilized the vector space where all the words of the lyrics were included [21] which was later on refined by Xia et al where the vector space included only those words that conveyed some emotions and was called s-VSM (sentiment-Vector Space Model). This led to dimension reduction and identifying what words are related to what sentiments.

V.II.II Emotion Recognition from Lyrics

The above described PAD value approach was used in lyric system which provided an aggregate score based on the PAD labels for all the words of the lyric. Meyers' system called Lyricator used this approach [22]. Lyricator extracted these emotions based on the summation of PAD values for all the words in the lyrics which fell into one of the four quadrants of the V-A model.

In order to utilize the capabilities of Machine Learning (ML) Xia and Chen made use of SVM (Support Vector Machine) for emotion identification. They had a collection of 2600 Chinese songs out of which 60% were labelled as light-hearted and the rest 40% as heavy-hearted. After training on the dataset their s-VSM feature set was more than 73% precise.

Hu and others used fuzzy clustering method to find out the emotion from the lyrics. Different clusters are obtained based on the words of the lyrics. These clusters are then ranked among themselves using grammatical information. This information gives weight and confidence via factors like sentence and word relationships and the cluster with the highest weight is considered the dominant emotion of the song. Though more efficient than Meyers' Lyricator [23], this approach assumes that each song exhibit only one emotion throughout its duration.

VI. CONTENT-BASED ANALYSIS

The traditional methods for deriving the information by manual annotation is bound to be swept off by the growth of the music libraries and repositories online and manual annotation in such cases would prove to be inefficient. In such scenario trying to recognize the emotion from the audio would be a better and a more refined approach. There are various acoustic features or attributes such as loudness, tempo and timbre which are known to affect the emotion of the music. Although multiple attributes are known to affect the emotion of the music, no single attribute is known to over weigh all the others.

Mion and De Poli attempted to find the most informative feature (in terms of mood) that could be extracted from the audio. They looked into feature selection system and showed initially on the set of one dimensional features that included intensity, spectral shape and other features [24]. They implemented SFS (Sequential Feature Selection) followed by PCA (Principal Component Analysis) to remove unnecessary features.

However, a limitation was that their study was focused on monophonic (single musical note) instrument classification, something which rarely occurs in a commercial music. Out of the tested features, roughness, attack time, peak sound level and notes per second were found to be most informative.

MacDorman and others used multiple features to determine arousal and valence values of the music. Features like sonogram, MFCC, fluctuation pattern and periodicity and spectral histogram were used. It was observed that these features gave better prediction of arousal values than valence.

Schmidt and others used multiple features. Individual features were used first and then used in combination to evaluate the performance of the both. Features like MFCCs, flux, centroid, roll-off, Chroma and octave-based spectral features were used [25].

A general trend was observed that the systems gave better prediction when multiple features were used in combination as opposed to individual features.

VI.I **Content-Based Systems**

The content-based analysis discussed earlier uses two types of model for emotion/mood identification. These are called as categorical and parametric models. As the name suggests, the former model is a classification model whereas the latter one is a regression model. We analyse the two models as follows.

VI.I.I **Categorical Models**

The publication by Li et al. used features like rhythm, pitch and timbre to train SVM model to classify music into one of the 13 categories. They manually labelled 499 clips of 30 seconds each having songs for variety of genres like classical, fusion, jazz with an accuracy of 45% [26].

Then came the improvement upon the above mentioned system where Lu, Liu and Zhang used features such as intensity, timbre and rhythm to train the Gaussian Mixture Model on the V-A model. 800 music clips of 20 second duration each were used and manually labelled to one of the four quadrants. Their system achieved an accuracy of 85% [27].

Tzanetakis was able to achieve a high accuracy of 61.5% by using only MFCC, spectral shape, roll off and centroid features where the classifier used was SVM [28].

Peeters in 2008 demonstrated his system which showed some improvement over Tzanetakis's system with an improved accuracy of 63.7% by using a larger feature base which included MFCCs, various Chroma as well as Spectral Crest. The classifier used was GMM (Gaussian Mixture Model) but to select the most informative features (40) the system employed Inertia Ratio Maximization with Feature Space Projection (IRMFSP) and performed LDA (Linear Discriminant Analysis) for dimensionality reduction [29].

In the next year Li and Cao demonstrated a system that performed with an accuracy of 65.7%. Their system used a super-vector of the low-level features, and implemented a GSV (Gaussian Support Vector) followed by SVM (Support Vector Regression) [30].

VI.I.II **Regression Models**

Schmidt and Han each started with V-A space and employed SVM for classification []. But the results were not up to the mark and unsatisfactory. While Schmidt obtained an accuracy of 50.2%, Han obtained an accuracy of 33% thus they moved to regression based methods where Han reformulated the problem using regression where the results obtained were mapped to the original mood categories. For this, GMM and SVR algorithms were used. Using GMM for regression and 11 quantized categories the obtained accuracy was around 95%.

Aiming at the results obtained through V-A coordinates obtained from the audio clips Yang and others used regression for mapping acoustic features into the 2-D space. D. Shrestha and D. Solomatine used an algorithm AdaBoost.RT [31] for regression and a ground truth V-A label was collected for all 195 clips. Features for this task were extracted using tools like Marsyas and PsySound, extracting a total of 114 features followed by PCA (Principal Component Analysis) for dimensionality reduction.

Eerola and others proposed a 3-D model for emotion prediction using only regression [32]. For this, they studied various regression techniques including PLS (Partial Least Squares) which takes into account the correlation between label dimensions.

Schmidt and others noticed that quantization of the V-A space by quadrants is inconsistent since the V-A labels collected were continuous in nature. Their team approached the problem using both SVR and MLR (Multiple Linear Regression).

VII. COMBINATION OF MULTIPLE DOMAINS

So far, it could be assessed that some features of any music could not be expressed via the audio itself and such aspects are determined by status of the song or audio clip in the society, the words used or as how the song is interpreted. Also, the acoustic features can give performance only up to a certain level. Hence, attempts were made to combine various annotation techniques and attempting to get the most accurate systems by the combination of features from multiple domains. Two such domains discussed here are tags and lyrics. Compared to other techniques, multi-modal techniques are still at a preliminary level.

VII.I **Combining Audio and Tags**

MIR community have focused on multi-modal approach of using tags with acoustic features of low-level since tagging techniques have been around for some time. Turnbull and others used web documents as well as social tags sources for tag collection and compared various algorithms like Calibrated Score Averaging, and RankBoost and used CAL500 dataset for audio analysis and found that the multimodal approach was able to perform better than approaches based on a single domain [34].

Bischoff and others also combined information from social tags and audio analysis for predicting the emotion. The social tags were collected from Last.fm platform and obtained 240 dimensional features including the popular MFCCs, Chroma and other spectral features. In order to classify social tags naïve Bayes classification algorithm was used and for feature classification SVM was used and combined them using a simple weighted approach [35].

VII.II **Combining Audio and Lyrics**

Yang and Lee's system used lyrics and audio features such as BPM (Beats per Minute) and 12 low level audio features. The dataset featured 145 30 second clips where each one was manually annotated [36]. There was strong indication of emotions according to the lyrics used but the performance was found to increase by a mere 2.1% only. The reason could be attributed to the use of a small database.

Yang and others also combined lyrics and audio features. The lyrics were analysed using BOW approach and 1240 Chinese pop songs were used where a 30 second clip from the middle of each song was extracted and were manually labelled using Thayer's VA model [37]. Firstly, the emotions were obtained using only the audio features and lyrics separately to classify arousal and valence using SVM and then merging the result to obtain a complete VA classification. The dataset was split into 80/20 training and testing sets and using only audio features an accuracy of 46.6% was obtained while using the combined approach the accuracy obtained was 57.1%.

Laurier and others also used audio and lyrical features to classify emotions in the VA space. The songs here unlike others were labelled using last.fm tags. The database consisted of 1000 songs and audio features like timbre, rhythm, tonal and temporal features were used. In order to extract lyrical features three approaches were used, LSA, LMD (Language Model Differences) and lyric similarity. LMD compares the difference in the frequency of the words used between language models of different mood categories. From this, 100 different words are obtained that are more significant resulting in higher performance. The performance improved for 'happy' and 'sad' moods by 5.1%. Since 'relaxed' and 'angry' predictions were already good enough, and the difference was very low hence they were not changed [38].

Hu and others also employed this combined approach. 18 emotion classes were formed based on social tags from last.fm and for refining these tags, BOW approach was used as well as cross checked by human judges. The size of dataset was 3000 songs which was relatively higher than their peers. Audio along with lyrics system was able to identify 13 out of 18 classes with high performance where audio alone performed best in emotions, happy, upbeat and desire whereas lyrics alone performed best in emotions, grief and exciting. The combined approach performed best under five categories i.e., calm, sad, anger, confident, earnest. The combined system showed improved performance in over half the cases where the standalone systems outperformed each other thus justifying the use of combined approaches.

VIII. CONCLUSION

Recognition of emotions remains a challenging task due to the underlying ambiguities in the way human express emotions. The research has advanced significantly since the inclusion of mood recognition as an MIR task since 2007 and the performance of automated systems have shown improvement using both single and multi-domain features and most accurate predictions have been obtained by employing machine learning algorithms using large feature set where the clips are of short length. But for further improvement, scientists from various domains (neurologists, psychologist) need to come together to understand the mood in the music as well as leading to better knowledge of human emotions in general. A joint effort of better understanding of human emotions would lead to improved automated systems.

This paper lays down a timeline based survey of what all has been done in the field of emotion recognition giving a basic outline of the major achievements and highlighting directions where further studies could reveal more information. As the need increases, the researchers are expected to collaborate in order improve the existing systems in the future.

REFERENCES

[1] C. C. Pratt, Music as the language of emotion. The Library of Congress, December 1950.

[2] Youngmoo E. Kim, Erik M. Schmidt, Raymond Migneco, Brandon G. Morton Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck, and Douglas Turnbull, "Music emotion recognition: a state of the art review" in ISMIR, 2010.

[3] Yang Liu, Yan Liu, Yu Zhao, and Kien A Hua, "What strikes the strings of your heart? Feature mining for music emotion analysis," IEEE Transactions on Affective Computing, Vol. 6, No. 3, pp. 247–260, 2015.

[4] Mathieu Barthet, Gyorgy Fazekas, and Mark Sandler "Music emotion recognition: From content- to contextbased models," Computer music modeling and retrieval, pp. 228–252, 2012.

[5] Alicja AWieczorkowska, Piotr Synak, and ZbigniewW Ra´s, "Multi-label classification of emotions in music," In Proc. of Intelligent Information Processing and Web Mining, Vol. 35, pp. 307–315, 2006.

[6] J. S. Downie, "The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research," Acoustical Science and Technology, Vol. 29, No. 4, pp. 247–255, 2008.

[7] X. Hu, J. Downie, C. Laurier, M. Bay, and A. Ehmann, "The 2007 MIREX audio mood classification task: Lessons learned," in Proc. of the Intl. Conf. on Music Information Retrieval, Philadelphia, PA, 2008.

[8] K. Hevner, "Experimental studies of the elements of expression in music," American Journal of Psychology, Vol. 48, No. 2, pp. 246–267, 1936.

[9] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement." Emotion, Vol. 8, pp. 494, 2008.

[10] E. Bigand, "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts," Cognition and Emotion, Vol. 19, No. 8, pp. 1113, 2005.

[11] U. Schimmack and R. Reisenzein, "Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation," Emotion, Vol. 2, No. 4, pp. 412, 2002.

[12] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multilabel classification of music into emotion," in Proc. Of the Intl. Conf. on Music Information Retrieval, Philadelphia, PA, 2008.

[13] F. Miller, M. Stiksel, and R. Jones, "Last.fm in numbers," Last.fm press material, February 2008.

[14] E. L. M. Law, L. von Ahn, R. B. Dannenberg, and M. Crawford, "TagATune: A game for music and sound annotation," in Proc. of the Intl. Conf. on Music Information Retrieval, Vienna, Austria, 2007.

[15] M. I. Mandel and D. P. W. Ellis, "A web-based game for collecting music metadata," in Proc. of the Intl. Conf. on Music Information Retrieval, Vienna, Austria, 2007, pp. 365–366.

[16] Y. Kim, E. Schmidt, and L. Emelle, "Moodswings: A collaborative game for music mood label collection," in Proc. Intl. Conf. on Music Information Retrieval, Philadelphia, PA, September 2008.

[17] O. Celma, P. Cano, and P. Herrera, "Search sounds: An audio crawler focused on weblogs," in Proc. of the Intl. Conf. on Music Information Retrieval, Victoria, Canada, 2006.

[18] P. Knees, T. Pohle, M. Schedl, D. Schnitzer, and K. Seyerlehner, a Document-Centered Approach to a Natural Language Music Search Engine. Springer Berlin / Heidelberg, 2008, pp. 627–631.

[19] A. Mehrabian and J. A. Russell, an Approach to Environmental Psychology. MIT Press, 1974.

[20] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW)," The NIMH Centre for the Study of Emotion and Attention, University of Florida, Tech. Rep., 1999.

[21] R. H. Chen, Z. L. Xu, Z. X. Zhang, and F. Z. Luo, "Content based music emotion analysis and recognition," in Proc. Of the Intl. Workshop on Computer Music and Audio Technology, 2006.

[22] O. C. Meyers, "A mood-based music classification and exploration system," Master's thesis, Massachusetts Institute of Technology, June 2007.

[23] Y. Hu, X. Chen, and D. Yang, "Lyric-based song emotion detection with affective lexicon and fuzzy clustering method," in Proc. of the Intl. Society for Music Information Conf., Kobe, Japan, 2009.

[24] L. Mion and G. D. Poli, "Score-independent audio features for description of music expression," IEEE Transactions on Audio, Speech and Language Processing, Vol. 16, No. 2, pp. 458–466, 2008.

[25] E. M. Schmidt, D. Turnbull, and Y. E. Kim, "Feature selection for content-based, time-varying musical emotion regression," in MIR '10: Proc. of the Intl. Conf. on Multimedia Information Retrieval, Philadelphia, PA, 2010, pp. 267–274.

[26] T. Li and M. Ogihara, "Detecting emotion in music," in Proc. of the Intl. Conf. on Music Information Retrieval, Baltimore, MD, October 2003.

[27] L. Lu, D. Liu, and H. J. Zhang, "Automatic mood detection and tracking of music audio signals," IEEE Transactions on Audio, Speech and Language Processing, Vol. 14, No. 1, pp. 5–18, 2006.

[28] G. Tzanetakis, "Marsyas submissions to MIREX 2007," MIREX 2007.

[29] G. Peeters, "A generic training and classification system for MIREX08 classification tasks: Audio music mood, audio genre, audio artist and audio tag," MIREX 2008.
[30] C. Cao and M. Li, "Thinkit's submissions for MIREX2009 audio music classification and similarity tasks." ISMIR, MIREX 2009.
[31] D. Shrestha and D. Solomatine, "Experiments with AdaBoost. RT, an improved boosting scheme for regression," Neural Computation, Vol. 18, No. 7, pp. 1678–1710, 2006.
[32] T. Eerola, O. Lartillot, and P. Toiviainen, "Prediction of multidimensional emotional ratings in music from audio using multivariate regression models," in Proc. of the Intl. Society for Music Information Conf., Kobe, Japan, 2009.
[33] E. M. Schmidt, D. Turnbull, and Y. E. Kim, "Feature selection for content-based, time-varying musical emotion regression," in MIR '10: Proc. of the Intl. Conf. on Multimedia Information Retrieval, Philadelphia, PA, 2010, pp. 267–274.
[34] D. Turnbull, L. Barrington, M.Yazdani, and G. Lanckriet, "Combining audio content and social context for semantic music discovery," ACM SIGIR, 2009.
[35] K. Bischoff, C. S. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo, "Music mood and theme classification-a hybrid approach," in Proc. of the Intl. Society for Music Information Retrieval Conf., Kobe, Japan, 2009.
[36] D. Yang and W. Lee, "Disambiguating music emotion using software agents," in Proc. of the Intl. Conf. on Music Information Retrieval. Barcelona, Spain: Universitat Pompeu Fabra, October 2004.
[37] Y.-H. Yang, Y.-C. Lin, H.-T. Cheng, I.-B. Liao, Y.-C. Ho, and H. Chen, Advances in Multimedia Information Processing - PCM 2008, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, December 2008, Ch. 8, pp. 70–79.
[38] C. Laurier, J. Grivolla, and P. Herrera, "Multimodal music mood classification using audio and lyrics," in Proc. of the Intl. Conf. on Machine Learning and Applications. Universitat Pompeu Fabra, 2008, pp. 1–6.

Authors Profile

Parichay Agrawal is pursing Bachelor of Technology from AKTU, India. His main research work focuses on Machine Learning.

Sahil Singh is pursing Bachelor of Technology from AKTU, India. His main research work focuses on Machine Learning.

Prashant Sharma is pursing Bachelor of Technology from AKTU, India. His main research work focuses on Machine Learning.

