

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

A Survey on Genre-Adaptive Semantic Computing and Audio-Based Modeling for Music Mood Annotation

Karandeepsingh Gill, Prof. G. K. Pakle,

PG Student, Department of Information Technology, Shri Guru Gobind Singhji Institute of Engineering and
Technology, Nanded, India

SGGSIE&T, Department of Information Technology, Shri Guru Gobind Singhji Institute of Engineering and
Technology, Nanded, India

ABSTRACT: The main point of music lies in its capability to express moods, so mood-based-music discovery and management are highly beneficial. Here we find whether taking genre into account is beneficial for automatic music mood annotation in terms of core effects valence, arousal, and tension, as well as several other mood scales. Novel techniques employing genre-adaptive semantic computing and audio-based modeling are proposed. A technique called the ACTwg employs genre-adaptive semantic computing of mood-related social tags, whereas ACTwg-SLPwg combines semantic computing and audio-based modeling, both in a genre-adaptive manner. In particular, improvements in the prediction rates are obtained for the valence dimension which is typically the most challenging core affect dimension for audio-based annotation. The specificity of genre categories is not crucial for the performance of ACTwg-SLPwg. It examined how genre information can be incorporated into music mood prediction using genre-adaptive semantic computing and genre-adaptive audio based modeling.

KEYWORDS: Music information retrieval, mood prediction, social tags, semantic computing, music genre, genre-adaptive.

I. INTRODUCTION

Music listening has been revolutionized in recent times with the shift to the digital age and the introduction of online music services. Online digital music sets include millions of tracks, and music streaming enables listeners to access, virtually, all the music ever recorded. Moreover, semantic metadata related to mood and genre has been shown to be amongst the most important ones for machine-based semantic music annotation or auto-tagging. It is easy to see why. On one hand, psychological studies have shown that music can be organized according to perceived and induced (elicited) emotions and music's ability to convey and affect moods is a key factor in explaining why music is culturally important. Using large data sources for music mood annotation has two advantages: 1) it enables operating on data that better represent large modern-day music catalogues and thus enable tapping into global characteristics of the relationship between music and emotion; and simultaneously, 2) it enables drawing information on this relationship at a more detailed level, by considering genre-specific aspects for example. While audio-based techniques have been efficient at predicting genres, mood prediction has remained more elusive, especially on the valence dimension, relating to the distinction between positive and negative emotions. Taking into account the genre-specificity of music moods may provide a way to alleviate this issue. Certain mood tags are more relevant to one genre than to another and audio-based mood annotation models trained on sets of tracks drawn from a particular genre give more accurate predictions within the corresponding genre than across genres.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

II. MOTIVATION

Employing audio-based techniques, semantic computing and genre-adaptively in music mood annotation have been beneficial in recent studies, while results have been boosted by using semantic computing or genre-adaptively in conjunction with audio-based techniques. However, it has not yet been investigated whether mood prediction performance could be boosted further by genre-adaptive semantic computing or by combining genre-adaptive semantic computing with audio-based techniques. Audio-based techniques have been efficient at predicting genres, mood prediction has remained more elusive, especially on the valence dimension, relating to the distinction between positive and negative emotions taking into account the genre-specificity of music moods may provide a way to alleviate this issue. Certain mood tags are more relevant to one genre than to another and audio-based mood annotation models trained on sets of tracks drawn from a particular genre give more accurate predictions within the corresponding genres than across genres.

III. LITERATURE SURVEY

- 1] **Mathieu Barthet, György Fazekas, and Mark Sandle, [Multidisciplinary Perspectives on Music Emotion Recognition: Implications for Content and Context-Based Models]:** In this paper, we first review state of the art studies on music and emotions from different disciplines including psychology, musicology and music information retrieval. Based on these studies, we then propose new insights to enhance automated music emotion recognition models. According to the categorical approach, emotions can be represented as a set of categories that are distinct from each other's. The dimensional models characterize emotions based on a small number of dimensions intended to correspond to the internal human representation of emotions. The appraisal approach was defined appraisal as a cognitive evaluation able to distinguish qualitatively among different emotions.
Advantages: 1) The high-level features (mode "meagerness" and key "clarity") is to enhance emotion recognition in a more robust way than low-level features.
2) The use of multiple modalities, most studies to date confirm that the strongest factors enabling emotion recognition are indeed related to the audio content.
Disadvantages: 1) The interpretability of the relationships between features and emotion components.
- 2] **Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, [MUSIC EMOTION RECOGNITION: A STATE OF THE ART REVIEW],** This paper surveys the state of the art in automatic emotion recognition in music. Music is oftentimes referred to as a "language of emotion", and it is natural for us to categorize music in terms of its emotional associations. The no. of features, such as harmony, timbre, interpretation, and lyrics affect emotion, and the mood of a piece may also change over its duration. But in developing automated systems to organize music in terms of emotional content, we are faced with a problem that oftentimes lacks a well-defined answer; there may be considerable disagreement regarding the perception and interpretation of the emotions of a song or ambiguity within the piece itself. When compared to other music information retrieval tasks (e.g., genre identification), the identification of musical mood is still in its early stages, though it has received increasing attention in recent years. In this paper we explore a wide range of research in music emotion recognition, particularly focusing on methods that use contextual text information (e.g., websites, tags, and lyrics) and content-based approaches, as well as systems combining multiple feature domains.
Advantages: 1) The performance of automated systems for music emotion recognition using a wide range of annotated and content-based features is increased.
2) It achieves predictions through large-scale machine learning algorithms.
Disadvantages: 1) Emotions in music change over time.
- 3] **P. Saari and T. Eerola, [Semantic Computing of Moods Based on Tags in Social Media of Music],** In this paper, propose a novel technique called Affective Circumplex Transformation (ACT) for representing the moods of music tracks in an interpretable and robust fashion based on semantic computing of social tags and research in emotion modeling. We validate the technique by predicting listener ratings of moods in music tracks and compare the results to prediction with the Vector Space Model (VSM), Singular Value Decomposition (SVD), Nonnegative

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Matrix Factorization (NMF), and Probabilistic Latent Semantic Analysis (PLSA). The results show that ACT consistently outperforms the baseline techniques, and its performance is robust to a low number of track-level mood tags. The results give validity and analytical insights for harnessing millions of music tracks and associated mood data available through social tags in application development.

Advantages: 1) The mood prediction with ACT is robust against the low number of tracklevelmood tags,2) Knowledge of moods is highly beneficial to semantic computing.

Disadvantages: 1) The possible discrepancy between track-level and clip-level moods, which may have reduced the prediction rates.

- 4] **M. Kaminskas and F. Ricci, [Contextual music information retrieval and recommendation: State of the art and challenges],** The basic idea is to retrieve and suggest music depending on the user's actual situation, for instance, emotional state, or any other contextual conditions that might influence the user's perception of music. Despite the high potential of such idea, the development of real-world applications that retrieve or recommend music depending on the user's context is still in its early stages. This survey illustrates various tools and techniques that can be used for addressing the research challenges posed by context-aware music retrieval and recommendation. This survey covers a broad range of topics, starting from classical music information retrieval (MIR) and recommender system (RS) techniques, and then focusing on context-aware music applications as well as the newer trends of affective and social computing applied to the music domain.

Advantages: 1) The advantage of annotation games over other human labor-based tagging methods is that the tagging cost is reduced

Disadvantages: 1) The difficulty of acquiring high-quality tags is the main reason why the potential of tags in music retrieval and recommendation.

- 5] **Y. Panagakakis, C. Kotropoulos, and G. R. Arce, [Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification],** Non-negative multilinear principal component analysis (NMPCA) is proposed for the unsupervised dimensionality reduction of the third-order tensors. The NMPCA maximizes the total tensor scatter while preserving the non-negativity of auditory representations. An algorithm for NMPCA is derived by exploiting the structure of the Grassmann manifold. The NMPCA is compared against three multilinear subspace analysis techniques, namely the non-negative tensor factorization, the high-order singular value decomposition, and the multilinear principal component analysis as well as their linear counterparts, i.e., the non-negative matrix factorization, the singular value decomposition, and the principal components analysis in extracting features that are subsequently classified by either support vector machine or nearest neighbor classifiers.

Advantages: 1) The best classification accuracy (84.3%) was obtained when NMPCA extracts features.2) To preserve the non-negativity of N-order tensor representations.

Disadvantages: 1) The multilinear dimensionality reduction techniques are unsupervised.

Sr. No.	Author paper ref. no.	Approach Technique	Dataset used and from where	Performance / Success ratio / Percentage
1	Mathieu Barthet, György Fazekas, and Mark Sandle, "Multidisciplinary Perspectives on Music Emotion Recognition: Implications for Content and Context-Based Models", in Proceedings of the 9 th International Symposium on Computer Music Modeling and	Machine Learning for Music Emotion Recognition	Two dataset used: 1) 72 tags from the CAL-500 dataset (genres, instruments, emotions, etc.) in category approach. 2) 4 quadrants of the Thayer-Russell AV space ("Exuberance", "Anxious/Frantic", "Depression", "Contentment")	The system achieved a performance increase of approximately 20% points (60.6%) in comparison with the system by Feng, Cheng and Yang (FCY1), proposed at MIREX 2009.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

	Retrieval (CMMR), 2012, pp. 492–507.			
2	Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, “MUSIC EMOTION RECOGNITION: A STATE OF THE ART REVIEW”, in Proceedings of the 11th International Conference of Music Information Retrieval (ISMIR). Citeseer, 2010, pp. 255–266.	Explore a wide range of research in music emotion recognition, particularly focusing on methods that use contextual text information (e.g., websites, tags, and lyrics) and content-based approaches, As well as systems combining multiple feature domains.	CAL500 data set This data set contains one song from 500 unique artists each of which have been manually annotated by a minimum of three no expert reviewers using a vocabulary of 174 tags, of which 18 relate to different emotions.	Audio + lyric approaches demonstrate the highest performance in recognizing 13 of the 18 mood classes.
3	P. Saari and T. Eerola, “Semantic Computing of Moods Based on Tags in Social Media of Music”, IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 10, pp. 2548–2560, 2014.	Vector Space Modeling, ACT of Mood Term Space. Mood Prediction of Tracks with ACT, Base line technique for mood prediction.	Large-scale mood data sets for music research, Gathering Vocabularies, Fetching Tags and Tracks from Last.fm,	ACT yields consistently high performance.
4	M. Kaminskas and F. Ricci, “Contextual music information retrieval and recommendation: State of the art and challenges,” Computer Science Review, vol. 6, no. 2, pp. 89–119, 2012.	Recommendation techniques—collaborative and Content-based filtering.	The Listen Game annotations were inferior to those of CAL500 dataset.	The automatic annotation model improved the performance.
5	Y. Panagakis, C. Kotropoulos, and G. R. Arce, “Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification,” IEEE Transactions on Audio, Speech, and Language Processing, vol.	Bio-Inspired Joint Acoustic And Modulation Frequency Representation of Music, Multilinear Subspace Analysis Techniques, Non-negative multilinear principal components analysis	The GTZAN and the ISMIR2004 Genre datasets demonstrate the superiority of NMPCA.	The NMPCA outperforms the other subspace analysis techniques either multilinear Or linear ones when it is used to extract features that are next classified by NN and SVM classifiers. The classification accuracy obtained when NMPCA selects auditory temporal modulations that are next classified by the SVM with an

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

18, no. 3, pp. 576–588, 2010.	(NMPCA) Algorithm, non-negative tensor factorization (NTF).		RBF kernel exceeds 70% even when 100 training samples are exploited.
-------------------------------	---	--	--

IV. EXISTING SYSTEM APPROACH

In the Existing system, the majority of research in music and mood has utilized either categorical (e.g., happiness, sadness and anger) or dimensional models of emotion. A well-known example of the latter is the effective circumplex which represents different emotions in the underlying dimensions of valence, distinguishing between positive and negative emotions. Semantic computing techniques, often based on Latent Semantic Analysis (LSA) have been employed to represent music moods based on tag data. Saari & Eero proposed an LSA-based technique called the Affective Circumplex Transformation (ACT) that yielded significant improvements over other techniques, as well as raw tags, at predicting listener ratings of the perceived core affects in music. The model training was carried out using social tags from Last. SLP involves projecting tracks to moods via a two-stage process, whereby a corpus of tracks is first mapped based on associated tags to a semantic space. Are trained between audio features and the semantic space. SLP outperformed conventional regression models trained to map audio features directly to listener ratings.

DISADVANTAGES OF EXISTING SYSTEM

- 1) The finding optimal “resolution” of genres and the fuzziness of genre categories are important problems.
- 2) In audio based techniques generating such ground-truth for large data sets in a controlled manner is often prohibitively laborious.

V. PROPOSED SYSTEM APPROACH

We develop new approaches for how genre information can be incorporated into music mood prediction using genre adaptive semantic computing and genre-adaptive audio based modelling. As the general baseline technique, SLP performed positive when compared to a state-of-the-art auto-tagging method. Proposes a technique that uses genre-adaptive semantic computing; it proposes a technique that employs both semantic computing and audio-based mood annotation in a genre-adaptive manner. It judges the effect of the specificity of genre categories for genre-adaptive mood annotation. The techniques are trained on a large set of social tag data and audio and evaluated for the prediction of listeners’ ratings of the perceived moods in a separate set of music tracks. The highest performing novel technique ACTwg- SLPwg models both the semantic mood space and audio-to mood relationship in a genre-adaptive manner. Moreover, audio-based genre inference for a novel track performed positively compared to tag-based inference, which has positive implications for applying the models to large unannotated datasets.

ADVANTAGES OF PROPOSED SYSTEM

- 1) This technique could be applied to other tasks, such as object recognition from images, video auto-tagging or multimedia retrieval.
- 2) Here context-adaptive semantic modeling can combine with context-adaptive content-based prediction could be beneficial.
- 3) Genre information can be incorporated into music mood prediction using genre adaptive semantic computing and genre-adaptive audio based modeling.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

VI. CONCLUSION

The highest performing novel technique ACTwg- SLPwg models both the semantic mood space and audio-to mood relationship in a genre-adaptive manner. Moreover, audio-based genre inference for a novel track performed favorably compared to tag-based inference, which has positive implications for applying the models to large unannotated datasets. The proposed techniques could be applied in future, such as object recognition from images, video auto-tagging, or multimedia retrieval, where context-adaptive semantic modeling combined with context-adaptive content-based prediction could be beneficial. Here the comparison with genre-adaptive models showed that taking into account the genre information in mood annotation shows consistent improvements.

REFERENCES

- [1] Mathieu Barthet, György Fazekas, and Mark Sandle, "Multidisciplinary Perspectives on Music Emotion Recognition: Implications for Content and Context-Based Models", in Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR), 2012, pp. 492–507.
- [2] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "MUSIC EMOTION RECOGNITION: A STATE OF THE ART REVIEW", in Proceedings of the 11th International Conference of Music Information Retrieval (ISMIR). Citeseer, 2010, pp. 255–266.
- [3] P. Saari and T. Eerola, "Semantic Computing of Moods Based on Tags in Social Media of Music", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 10, pp. 2548–2560, 2014.
- [4] M. Kaminskas and F. Ricci, "Contextual music information retrieval and recommendation: State of the art and challenges," Computer Science Review, vol. 6, no. 2, pp. 89–119, 2012.
- [5] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 3, pp. 576–588, 2010.