Towards the Development of Music Mood Classification of Original Pilipino Music (OPM) Songs Based on Audio and Lyrics Keyword

Mideth B. Abisado
National University, Manila

Mardyon B. Yongson
Technological University of the Philippines, Manila

MA. Ian P. De Los Trinos
Technological University of the Philippines, Manila

ABSTRACT

This paper presents music mood classification of Original Pilipino Music (OPM) songs, particularly Filipino songs using audio and lyrics information. The song's mood is expressed utilizing musical features, but a relevant part also seems to be conveyed by the keywords to its lyrics. The study evaluates with the help of two music teachers and music analysts each factor independently. It explores the possibility of combining both, using Natural Language Processing and Music Information Retrieval techniques. It shows that standard separation-based strategies and Latent Semantic Analysis can group the verses essentially superior to random. Yet, the exhibition is still very substandard compared to that of sound-based systems. The study presents a technique dependent on contrasts between language models that gives performances closer to sound-based classifiers—in addition, interwinding this in a multimodal framework, which is audio and text. It permits an improvement in the general execution. We exhibit that verses and sound data are corresponding and can be joined to improve an ordered framework.

CCS CONCEPTS

• Human-centered computing; **• Applied computing**; **• Arts and humanities**; **• Sound and music computing**; **• Modeling**;

ACM Reference Format:

Mideth B. Abisado, Mardyon B. Yongson, and MA. Ian P. De Los Trinos. 2021. Towards the Development of Music Mood Classification of Original Pilipino Music (OPM) Songs Based on Audio and Lyrics Keyword. In *2021 5th International Conference on E-Society, E-Education and E-Technology (ICSET 2021), August 21–23, 2021, Taipei, Taiwan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3485768.3485786

1 INTRODUCTION

The rapid development of music gushing administrations has led clients to approach progressively huge indexes of music. Because of progressions in innovation, music order and proposal have increased expanded ubiquity in the music industry. In this unique situation, assessment examination has an important job. It alludes to using characteristic language handling and content examination to distinguish what the creator implied or needed to pass on [1]. The utilizations of feeling investigation are many: archives, blogs, or recordings. One specific source where intrigue has been developing is music. Many existing music grouping frameworks use AI algorithms to arrange music by music analyst likeness [2], feeling [3], or sort [4]. Individuals might need to tune in to music dependent on their state of mind. Mental examinations [5] have demonstrated that tuning in to music has an upgrading impact on the audience members' social attachment, emotional state, and mindset. Music state of mind acknowledgment is the procedure wherein the feelings of a melodic piece are recognized through different methods, including investigating sound and musical content [6]. Most of the examination on music grouping depends on highlights acquired by good signs [7, 8].

Nonetheless, as appeared in [9], the semantic (verses) and consonant (tunes) data are prepared freely by the mind, in any event, when these data sources are firmly identified with one another. In this way, the investigation of verses alone as a wellspring of data can be applied in music characterization. It is a fascinating issue, and it has not been broadly investigated in writing.

1.1 Objectives

This research is centered around utilizing word-level highlights for the state of music and mood classification. This research examines which highlights extraction technique is the most encouraging and productive for the state of mind grouping and fabricates a classifier that can foresee whether a melody is happy or sad depending on its audio and music features. This classifier could be additionally enclosed by a music proposal framework to choose music by slant in various social settings.

2 RELATED STUDIES

The research utilizes existing techniques, including a similar errand of characterization and expectation. A wide cluster of potential arrangements, particularly on the utilization of Artificial Intelligence, notably machine algorithms, can be browsed. Accordingly, they must be cautiously chosen to yield the best results. The succeeding sections talk about this accurateness independently.

Elevated levels of characterization exactness have been accomplished utilizing AI procedures. One study [10] specifically yielded correctnesses of 90% or higher. Fujinaga and McKay used kNN and neural systems. They put an incredible incentive on elevated level highlights. Also, utilizing a more extensive arrangement of 'significant highlights' gives better outcomes. Another study [11–13] concentrated on mood and pitch. They used a Gaussian Classifier

ICSET 2021, August 21–23, 2021, Taipei, Taiwan

ACM ISBN 978-1-4503-9015-6/21/08…$15.00
https://doi.org/10.1145/3485768.3485786

just as kNN on 10-overlay Cross-Validation. In [14], the examination gives an inside and out survey into musical highlights and the acknowledgment of instruments. In addition, the mechanisms inside the investigation were likewise connected to particular classes. Fuhrmann surveyed the melodic highlights and its suggestions on the sound created for each instrument in the analysis. The connection is again made between comparative-sounding devices and the class they have a place with. Also, Fuhrmann utilized Weka to run classifiers and researched Spectral highlights and MFCC, among others, in the examination. Xu et al. [15, 16] used the Support Vector Machine to characterize a lot of 60 tunes. They utilized a 'multi-layer approach wherein tunes are first arranged into pop/old style and shake/jazz before they are put during a time Support Vector Machine, or another 'layer,' which would sort the tunes between them the two classifications of the parent hub. For instance, the subsequent layer will characterize songs among pop and old-style on the off chance it was arranged into pop/old style in the first SVM layer.

The analysts likewise tried their informational index utilizing different classifiers to survey the nature of their multi-layered SVM approach. Contrasted with the Hidden Markov Model, kNN, and Gaussian blend model, the multi-layered SVM approach had a minimal measure of blunder at 6.86%. Neumayer and Rauber [17, 18] consolidated content and low-level sound highlights for sort grouping. The examination includes preliminaries of order utilizing sound just, a message just, and a mix of both. The musical highlights incorporate vacillation designs (cadence design), mood histograms, and Statistical Spectrum Descriptors. Out of all preliminaries directed, the best arrangement precision results from a mix of verses, content, and musicality designs. A rhythm-driven study demonstrated the significance of beat is significant in exact sort arrangement [19]. In light of their discoveries, rhythm alone to arrive at 80% arrangement precision bolsters their case. Notwithstanding, they used just one out of numerous highlights that add to musicality. One can't consequently accept that accuracy of an order will improve if more descriptors are included.

3 METHODS AND RESULTS

3.1 Dataset

The vast majority of the open-source music datasets are constrained to sound highlights. For instance, a dataset for music data recovery is the free Million Song Dataset that contains sound highlights and metadata of a million music tracks. The absence of open verses datasets is because of the way that these writings are copyrighted material. A generally utilized verses dataset is discharged by musixmatch.com, which gives verses as a prepared pack of words. Be that as it may, the decision of gaining the verses in a natural configuration over the musixmatch.com dataset was important for looking at changed element extraction and preprocessing steps.

3.2 Mood Labeling

Mood and song details (artist and title) were automatically collected and classified with the help of a music teacher and music analyst. The dataset contains song-level tags provided by music analysts for more than 200 songs. Custom code was written to download the details of songs labeled as happy and sad. Two hundred songs' details have been downloaded, 100 songs classified as sad, and 100 songs classified as happy with the help of music teacher and analyst.

3.3 Lyrics Collection

Using details from musixmatch and originalpinoy-lyrics.blogspot.com, a custom content management system for original Pinoy music (OPM) video and lyric collection. The URL is also included in the dataset. Aside from the lyrics.

3.4 Text Processing

The collected texts are not the ones that will be analyzed but have to be changed. Preprocessing method plays a significant role in text mining techniques and applications.

This stage divides the text into relevant units such as individual words (tokens) or sequences of words (n-grams). In particular, tokenization refers to the task of segmenting a text into individual words or word-like units. The RegexpTokenizer function from the NLTK library is used for tokenization, which removes all the non-alphanumeric characters.

It is worth highlighting the importance of stop-words removal. Stop-words should be removed from a text because they make it look heavier and less critical for analysts. Removing stop words reduces the dimensionality of term space. However, this work doesn't remove the stop-words. This can be motivated by the fact that the stop-word set includes words like" hindi na, " huwag, wag, " wag, " di, wala na," etc. which may prove of some significance when training the model.

3.5 Classifier

Given the dataset composed of feature vectors, the goal is to train a classifier to predict whether a song's lyrics are happy or sad and the Naive Bayes classifier provided by the NLTK library. It uses a Bernoulli model allowing only binary features, other classifiers provided by the sci-kit-learn library are analyzed: linear Support Vector Machine (SVM) using Stochastic Gradient Descent, non-linear SVM where the value of γ and C are found by grid-search, Bernoulli Naive Bayes classifier, Multinomial Naive Bayes classifier, and Random Forest classifier.

3.6 Metrics

The chosen metrics for answering RQ1 are accuracy, precision, recall, and F- measure of classifiers splitting the dataset in training (80%) and testing (20%) set in Weka or RapidMinder. The accuracy is the proportion of correctly classified items:

Accuracy = T P +T N TP+TN+FP+FN

The F-measure the harmonic mean of precision and recall:

F − measure = 2 ∗ precision∗recall

where and

precision+recall

P recision = T P TP+FP

Recall= TP TP+FN

(TP = number of true positives, FP = number of false positives, TN = number of true negatives, and FN = number of false negatives)

To answer to RQ2, two metrics are evaluated:

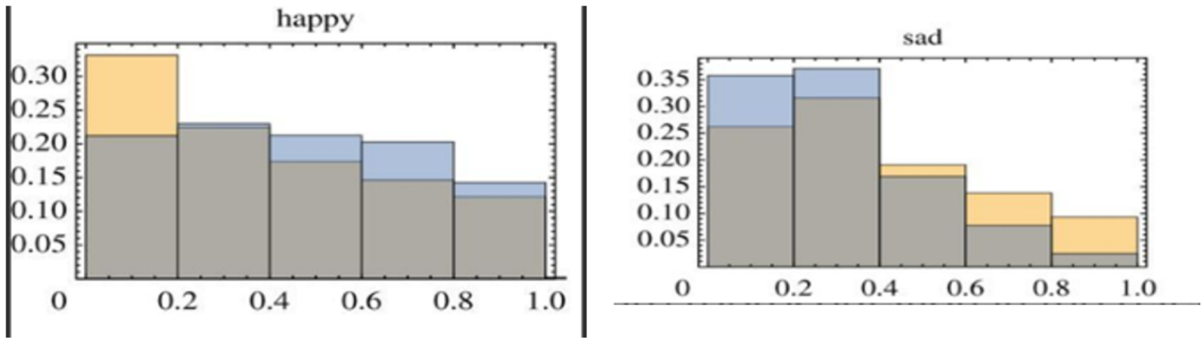

Figure 1: Classification Visualization

■ The accuracy was computed by 10-fold cross-validation. As defined in [20], cross-validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model. The other is used to validate the model. The division process can be repeated k times, using different subsets of the data. So, the idea of cross-validation is to estimate how well the current dataset can predict an output value for any given input instance.

■ AUC, or Area Under Curve, is another metric for binary classification. It's probably the second most popular one, after accuracy. While accuracy deals with getting the class label right or not, the AUC considers the uncertainty value about the answer quantified by a classifier. So, a threshold is needed: the simplest one is 0.5, but the advantage of AUC is that it considers all possible points. Various entries result in different true positive/false-positive rates. As you decrease the threshold, you get more true positives and more false positives. A score for a perfect classifier would be 1.

Figure 1 reflects different successful (blue) and unsuccessful (yellow) songs through the confusion matrix.

The results obtained with this approach, one could imagine a music classification through MIR and Librosa. Some interesting results confirm that lyrics and tunes information is processed independently by the brain leading to a different perception of a song. For example, listening to the music of "WAG KA NG UMIYAK" by KZ Tandigan seems to be a sad song, thanks to its rhythm. This infectiously catchy tune contains very sad lyrics dealing with heartbreak and the pain of letting someone go who you wish you could have back. This lyrical dissonance is confirmed by the implemented model that classifies this song as sad with a probability of 97%.

4 CONCLUSIONS

This study observes the following patterns: (a)Successful songs are happier than average songs, (b) Successful songs have a brighter timbre than average songs, (c) Successful songs are less sad than average songs,(d) Successful, happy songs are more party-like than average songs,(e) Successful sad songs are less relaxed than average songs,(f) Successful, happy songs are more danceable than average songs. Furthermore, the dataset can be augmented by crowd-sourcing: this classifier will be wrapped in a web application where users can test it and confirm the predictions and be able to see this pattern,

In this way, the number of mood-labeled lyrics can increase. Finally, further experimentations will be conducted adding more moods, like angry, tense, etc.

5 FUTURE WORK

This study is ongoing research. Further expansion of dataset and implementation and testing and evaluation of other algorithms shall be conducted. A mobile application and a browser plug-in will be developed when optimized results are met, implementing the classification of sad and happy songs.

REFERENCES

[1] Emil Ian V. Ascalon, Rafael Cabredo, Lyric-Based Music Mood Recognition, DLSU Research Congress 2015, 2015
[2] Tao Li and Mitsunori Ogihara. Music artist style identification by semi-supervised learning from both lyrics and content.
[3] Lie Lu, Dan Liu, and Hong-Jiang Zhang. Automatic mood detection and tracking of music audio signals. IEEE Transactions on audio, speech, and language processing, 14(1):5–18, 2006.
[4] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. Speech and Audio Processing, IEEE transactions on, 10(5):293–302, 2002.
[5] Thomas Sch äfer, Peter Sedlmeier, Christine St ädtler, and David Huron. The psychological functions of music listening. Frontiers in psychology, 4, 2013.
[6] Lie Lu, Dan Liu, and Hong-Jiang Zhang. Automatic mood detection and tracking of music audio signals. IEEE Transactions on audio, speech, and language processing, 14(1):5–18, 2006.
[7] M. F. McKinney and J. Breebaart, Features for Audio and Music Classification ISMIR, vol. 3, pp. 151–158, 2003.
[8] R. Mihalcea and C. Strapparava, "Lyrics, Music, and Emotions," In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 590–599, 2012.
[9] Porter, Martin F." An algorithm for suffix stripping." Program 14.3 (1980): 130-137
[10] Paice, Chris D. "Another Stemmer" ACM SIGIR Forum 24.3 (1990): 56-61.
[11] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. Speech and Audio Processing, IEEE transactions on, 10(5):293–302, 2002.
[12] Fuhrmann, F. (2012). Automatic musical instrument recognition from polyphonic music audio signals (Doctoral dissertation, Ph.D. thesis, Universitat Pompeu Fabra).
[13] J. Potvin, P. Soriano, M. Vallee "Generating trading rules on the stock markets with genetic programming
[14] M. Zaanen and P. Kanters, "Automatic Mood Classification using tf*idf based on Lyrics," in J. Stephen Downie and Remco C. Veltkamp, editors, 11th International Society for Music Information and Retrieval Conference, 2010.
[15] D. Yang and W-S Lee, "Music Emotion Identification from Lyrics," Eleventh IEEE Int. Symposium of Multimedia, 2009.

[16] 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.
[17] J. A. Russell, "A circumspect model of affect," Journal of Psychology and Social Psychology, vol. 39, no. 6, p. 1161, 1980.
[18] J. A. Russell, "Core affect and the psychological construction of emotion," Psychol. Review, 110, 1, 145–172. , 2003.
[19] D. C. Montgomery, G. C. Runger and N. F. Hubele, "Engineering Statistics," Wiley, 1998.
[20] K. Krippendorff, "Content Analysis: An Introduction to its Methodology," 2nd edition, chapter 11. Sage, Thousand Oaks, CA, 2004.