A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music

Shyamala Doraisamy **Shahram Golzari** **Noris Mohd. Norowi**
Md. Nasir B Sulaiman **Nur Izura Udzir**
Faculty of Computer Science and Information Technology
University Putra Malaysia

ABSTRACT

Machine learning techniques for automated musical genre classification is currently widely studied. With large collections of digital musical files, one approach to classification is to classify by musical genres such as pop, rock and classical in Western music. Beat, pitch and temporal related features are extracted from audio signals and various machine learning algorithms are applied for classification. Features that resulted in better classification accuracies for Traditional Malay Music (TMM), in comparison to western music, in a previous study were beat related features. However, only the J48 classifier was used and in this study we perform a more comprehensive investigation on improving the classification of TMM. In addition, feature selection was performed for dimensionality reduction. Classification accuracies using classifiers of varying paradigms on a dataset comprising ten TMM genres were obtained. Results identify potentially useful classifiers and show the impact of adding a feature selection phase for TMM genre classification.

1. INTRODUCTION

Interest on music information retrieval systems for the storage, retrieval and classification of large collections of digital musical files has grown in recent years. Metadata such as filename, author, file size, date and genres are commonly used to classify and retrieve these documents. Such manual classification is highly labour-intensive and costly both in terms of time and money [1].

An automatic classification system that is able to analyse and extract implicit knowledge of the musical files is therefore highly sought. One approach to musical classification that is currently being widely studied is classification by musical genres. Musical genres are labels created and used by humans for categorizing and describing music [2].

Examples of a few Western musical genres are such as Pop, Rock, Hip-hop, and Classical. Several systems for automated genre classification and retrieval of musical files have been researched and developed [2,3]. However, most of these studies were conducted using only western dataset and we focus on non-Western musical genres, and more specifically Traditional Malay Music (TMM).

TMM encompasses all traditional music from Malaysia, both West Malaysia (on the Peninsular) and the states on the Borneo Island (Sabah and Sarawak) [4]. Genre examples include Dikir Barat, Joget, Wayang Kulit, Gamelan, Etnik Sabah and Inang. In general, these musical genres have a strong sense of rhythm, partly due to the fact that TMM is traditionally played by ear as opposed to reading from written musical scores. Having the beat or rhythm clearly audible helps when the musical piece is being passed down orally through generations in the villages such as having clear gong hits. The significance of beat features for TMM genre classification in comparison to Western musical genres was investigated in a previous study using a two-phase methodology – feature extraction and classification [5]. In this paper we study the impact of adding a feature selection phase for TMM genre classification.

Feature extraction is a process where a segment of an audio is characterized into a compact numerical representation. Due to the high dimensionality of these feature sets, feature selection can be performed to reduce the dimensionality of the data as a preprocessing step prior to classification. With audio data, several studies have investigated the significance of this phase and the performance of several feature selection methods [21]. A comprehensive discussion on feature selection is available in Saeys et. al [6].

A large number of the studies performed on music classification have looked more closely at the feature extraction step, and in this study we investigate the classification phase more closely. The rest of this paper is organised as follows: Section 2 presents background information about music genre classification and TMM. An overview of feature selection is presented in Section 3. Section 4 presents the experimental framework and results are discussed in Section 5.

2. BACKGROUND

2.1. Music Genre Classification

Digital audio in general is categorized as speech, music and noise. Wold, et al[3] analyse and compare audio features such as rhythms, pitch, duration, loudness and instrument identification to classify various groups of audio such as speech, gender, animal sounds and sound effects. However, music classification was not emphasized in their study.

There has been great interest of classification of audio based on musical genres in recent years [2, 7,8]. Tzanetakis and Cook [2] and Aucoturier and Pachet[8] categorized audio features into three categories; timbral related features, rhythmic related features, and pitch related features. Audio data that are to be classified cannot be represented as raw audio data, such as samples of amplitude value of digital audio signals. Hence, some form of parameterisation is required. Parameterisation of audio data is based on audio analysis, which can be done using several methods such as Fourier transform, wavelet transform, statistical methods, etc. [9].

Timbral related features are based on the Short Time Fourier Transform (STFT), which is used to determine the phase content of short local sections in a signal as it changes over time. The features are used in music-speech discrimination and speech recognition. Examples of timbral related features are such as spectral centroid, spectral roll off and time domain zero crossing, which measure the spectral shape, the changes in spectral shape and the noisiness of a signal respectively. Another feature, Mel-Frequency Cepstral Coefficients (MFCC), is also based on the STFT, but is typically used to provide compact representation of the spectral envelope, especially in speech recognition. Tzanetakis and Cook [2] provide a detailed discussion on timbral related features and the account of their experiment.

Beat features analyses the signals that calculate the rhythmic structure of music based on their wavelet transform [2]. It involves time-frequency analysis, which is useful to music classification as its algorithm is similar to human hearing. The main beat can be defined as the regular periodic sequence of pulses corresponding to where a human would tap his foot whilst listening to music. Although achievable, extraction of beat features is very difficult. Whilst it is trivial for human to do so to a music, it is not so with machines. Kosina[10] and Dixon[11] give good overviews on beat tracking methods. Li and Tzanetakis[12] investigate the effects of different feature set combinations for optimum classification performance. Features incorporated in the study were: FFT, MFCC, Pitch and Beat. Although suitable feature set combinations from this study was obtained, it was also suggested that they might not be generic to all genres but applicable only to western genres that was used in their study.

With classification, classifiers vary in terms of robustness, speed, memory usage and complexity. Several studies investigate the use of existing classifiers for musical genre classification [2,3,12]. For instance, OneR is a primitive form of classifier as it produces simple rule based on one attribute only, but it is useful in determining a baseline performance as a benchmark for other learning schemes [13]. Emphasis on the importance of understanding different classifiers is also discussed at length by [7,14].

2.2. Traditional Malay Music

Traditional Malay music is mainly derivative, influenced by the initial overall Indian and Middle Eastern music during the trade era and later from colonial powers such as Thailand, Indonesia, Portuguese and British who introduced their own culture including dance and music. A thorough overview on the origin and history of TMM can be found in [17]. The taxonomy of TMM depends on the nature of the theatre forms they serve and their instrumentations. Categorization of TMM genres has been studied extensively by Ang[18]. Music of these genres is usually disseminated non-commercially, usually performed by persons who are not highly trained musical specialists, undergoes change arising from creative impulses and exists in many forms. The musical ensembles usually include gendangs or drums that are used to provide constant rhythmic beat of the songs and gongs to mark the end of a temporal cycle at specific part of the song [19].

One common attribute that is shared by most TMM genres is that they are generally repetitive in nature and exist in 'gongan'-like cycle. 'Gongan' is defined as a temporal cycle marked internally at specific points by specific gongs and at the end by the lowest-pitched gong of an ensemble [17]. It is an important structural function as it divides the musical pieces into temporal sections. Once every measure has been played, musicians continue playing in a looping motion by repeating the cycle from the beginning again until one of the lead percussionists signals the end of the song by varying their rhythms noticeably. Traditional Malay music does not have a chorus that plays differently than other parts of the songs, which is the usual occurrence in western music. Its repetitiveness and constant rhythms are two aspects that are taken into account to facilitate classification by genre later.

Very little study has been conducted on automatic traditional Malay music genre classification in the literature. Norowi, et al[20] study the effects of various factors and audio feature set combinations towards the classification of TMM genres. Results from experiments conducted in several phases show that factors such as dataset size, track length and location¸ together with various combinations of audio feature sets comprising Short Time Fourier Transform (STFT), Mel-Frequency Cepstral Coefficients (MFCCs) and Beat Features affect classification. Based on parameters optimized for TMM genres, classification performances were evaluated against three groups of human subjects:

experts, trained and untrained. Based on the result of this study performances of both machine and human were shown to be comparable. However, only the J48 classifier was used with 66.3% classification accuracy [5]. In this study, we assess the practical usefulness of a wide range of classifiers and identify potential classifiers that would improve the performance of TMM genre classification. We confine our study to the classifiers within WEKA which is discussed further in section 4.2.

3. FEATURE SELECTION

Feature selection is the process of removing features from the data set that are irrelevant with respect to the task that is to be performed. Feature selection can be extremely useful in reducing the dimensionality of the data to be processed by the classifier, reducing execution time and improving predictive accuracy (inclusion of irrelevant features can introduce noise into the data, thus obscuring relevant features). It is worth noting that even though some machine learning algorithms perform some degree of feature selection themselves (such as classification trees), feature space reduction can be useful even for these algorithms. Reducing the dimensionality of the data reduces the size of the hypothesis space and thus results in faster execution time.

In general, feature selection techniques can be split into two categories - filter methods and wrapper methods. Wrapper methods generally result in better performance than filter methods because the feature selection process is optimized for the classification algorithm to be used. However, they are generally far too expensive to be used if the number of features is large because each feature set considered must be evaluated with the trained classifier. For this reason, wrapper methods will not be considered in this study. Filter methods are much faster than wrapper methods and therefore are better suited to high dimensional data sets. Diverse feature ranking and feature selection techniques have been proposed in the machine learning literature, Such as:

- Correlation-based Feature Selection (CFS) [21]
- Principal Component Analysis (PCA) [21]
- Gain Ratio (GR) attribute evaluation [21]
- Chi-square Feature Evaluation [21]
- Support Vector Machine Feature Elimination (SVM-RFE) [22]

Some of these methods does not perform feature selection but only feature ranking, they are usually combined with another method when one needs to find out the appropriate number of attributes. Forward selection, backward elimination, bi-directional search, best-first search [13], genetic search [23], and other methods are often used on this task.

Fiebrink et. al [23] investigated the significance of the addition of feature selection with classification of Western musical genres. The results showed an almost similar classification accuracy using forward selection and PCA, a wrapper and filter method respectively. Classification utilized a fraction of time with PCA. We evaluate several filter methods for TMM genre classification in this study to achieve to our purpose: choose the best combination of feature selection and classification to classify the TMM genre.

4. EXPERIMENTS

The aim of this study is to investigate the impact of the addition of feature selection towards TMM classification. For this purpose some experiments were designed and conducted and explained in this section.

4.1. Data Set

The data collection and pre-processing stages of this study are described in these following sub-sections.

4.1.1. Data Collection

Ten TMM genres were involved in this study. The breakdown for each genre and its number of musical files are listed in Table 1. A relatively small dataset was used in this experiment due to the difficulty in obtaining digital files of TMM, as traditional Malay musical culture is fast corroding with little preservation in digital format. Whilst it was much easier to obtain dataset for western music, the number was also kept small to match the size of TMM dataset.

NO	Genre	Number
1	Dikir Barat	31
2	Etnik Sabah	12
3	Gamelan	23
4	Ghazal	17
5	Inang	10
6	Joget	15
7	Keroncong	43
8	Tumbuk Kalang	13
9	Wayang Kulit	17
10	Zapin	10

Table 1. Overall number of musical files for each genre

Musical files for this experiment were obtained from the Malaysia Arts Academy, Sultan Salahuddin Abdul Aziz Shah's Cultural and Arts Centre at Universiti Putra Malaysia, Student's Cultural Centre at Universiti Malaya and also personal collections of audio CDs from many individuals. The dataset became available in both digital and analog format. Quite a number of musical data for TMM genres were in analog format and were digitized manually. All of the digital music files were then converted into wav files; the only audio format supported by the existing feature extraction tool used at the time of study. The whole dataset was later trimmed to specific length and location in the file by executing certain audio commands through batch processing before extraction began.

4.2. Genre Classification Component

This section discusses feature extraction and classification using Musical Research System for Analysis ad Synthesis (MARSYAS) [2] and Waikato Environment for Knowledge Analysis (WEKA) [13]. We use MARSYAS for feature extraction and WEKA for feature selection and classification.

4.2.1. Feature Extraction

The features were extracted from the music files through MARSYAS-0.2.2; a free framework that enables the evaluation of computer audition applications. MARSYAS is a semi-automatic music classification system that is developed as an alternative solution for the existing audio tools that are incapable of handling the increasing amount of computer data [2]. It enables the three feature sets for representing the timbral texture, rhythmic content and pitch content of the music signals and uses trained statistical pattern recognition classifiers for evaluation. The feature extractor will produce numerical outputs in the form of Attribute Related File Format (ARFF) files. In this study we extracted STFT + MFCC + Beat Feature because this combination of features had been achieved best accuracy for TMM genre classification in [5] and also includes the complete set of features (73 features).

Scenario	Description
S1	**No Feature selection**
S2	**Correlation-based Feature Selection with best first search strategy**
S3	**Correlation-based Feature Selection with genetic search strategy**
S4	**Correlation-based Feature Selection with greedy stepwise search strategy**
S5	**Principal Component Analysis**
S6	**Chi-square Feature Evaluation**
S7	**Gain Ratio Feature Evaluation**
S8	**SVM based Feature Evaluation**

Table 2. Description of scenarios

4.2.2 Feature Selection

We used seven feature selection methods in the experiments of study. There are eight scenarios at the experiments in which one scenario does not incorporate a feature selection steps. The description of scenario is shown in Table 2.

4.2.3 Classification

To compare the performance of classification algorithms, the list of classifiers chosen included a wide range of paradigms. The code written was based on the WEKA data mining package and the default parameters used for each algorithm. All experiments were carried out using a ten-fold cross validation approach and to control the validity of experiments. The list of classifiers and results of the experiments are shown in Table 3. These include AIRS (a classifier based on the Artificial Immune System (AIS) paradigm [25,26]), Bagging, Bayesian Network, Cart, Conjunctive rule learner (Conj-Rules), Decision Stump, Decision Table, IB1, J48 (an implementation of C4.5), Kstar, Logistic, LogitBoost, Multi-layer neural network with back propagation (MLP), Naïve Bayesian, Nbtree, PART (a decision list [27]), RBF Network, and SMO (a support vector machine [28]).

Of these classifiers, AIRS is discussed a little further. It's performance for musical genre classification has not been widely investigated. AIS is a computational method inspired by the biology immune system. It is progressing slowly and steadily as a new branch of computational intelligence and soft computing [29]. One of the AIS based algorithms is the Artificial Immune Recognition System (AIRS). AIRS is a supervised immune-inspired classification system capable of assigning data items unseen during training to one of any number of classes based on previous training experience. AIRS is probably the first and best known AIS for classification, having been developed in 2001. [29]. This study also investigates the performance of this algorithm for musical genre classification.

5. RESULTS

Table 5 lists the classifiers that obtained highest classification accuracies for each of the described scenarios. The descriptive details of the classification are shown in Table 4.

The highest accuracy was obtained with MLP using the Correlation-based Features Selection and a genetic search strategy. Although none of the evaluated classifiers and feature selection methods provided us with a combination that outperforms a particular combination, useful knowledge was gained regarding combinations that do not contribute significantly to the task. The functional-based classifiers, MLP and SMO prove to be superior to the other classifiers. The combination of MLP and Correlation-based Feature Selection with genetic search strategies has achieved best accuracy of 88.6%.

	min	max	Avg
S1	32.64	86	68.13
S2	33.68	87	71.96
S3	33.68	88.6	71.36
S4	33.68	87	72.25
S5	35.23	83.42	60.88
S6	31.6	87.05	71.45
S7	31.6	87.05	72.37
S8	31.6	84.46	69.89

Table 4. Descriptive statistics of scenarios

Scenario	Max Accuracy (%)	Classifier
S1	86	MLP
S2	87	MLP
S3	88.6	MLP
S4	87	MLP
S5	83.42	SMO
S6	87.05	AIRS-SMO
S7	87.05	AIRS
S8	84.46	SMO

Table 5. Max accuracy achieved by classifiers in each scenario

30% is obtained. However, the use of PCA did not still improve the performance of AIRS.

6. CONCLUSION

A comprehensive study evaluating the performance of a wide range of classifiers for automating TMM genre classification was performed. Features that were found to be useful for TMM genre classification in comparison to Western genres continued to be used in this study. Results obtained clearly identify potentially useful classifiers -- the functional-based classifiers MLP and SMO. The impact of including a feature selection phase for TMM genre classification was investigated. The results show that the addition of this phase did improve the performance of most classifiers by at least 1%. This addition however improved the performance of the immune-based classifier, AIRS very much as discussed above. Future work will include further experiments to investigate these findings on improved musical genre classification with AIRS and a comparative study to Western Musical genres.

	Accuracy (%)								
Classifier	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**Avg**
AIRS	51.81	84.97	82.90	84.97	59.10	87.05	87.05	83.41	77.66
Bagging	73.57	74.61	75.13	75.65	61.66	75.13	76.68	68.91	72.67
Bayesian Network	78.24	78.76	77.20	78.76	49.22	78.24	80.83	75.13	74.55
Cart	58.55	60.10	62.18	63.73	60	60.62	61.67	61.14	61
Conj-Rules	32.64	33.68	33.68	33.68	35.23	31.60	31.60	31.60	32.96
Decision Stump	33.68	33.68	33.68	33.68	35.75	33.69	33.68	33.68	33.94
Decision Table	60	53.36	56.48	53.37	45.60	51.81	52.85	55.96	53.68
IB1	75.65	84.45	82.38	84.46	67.36	84.97	84.97	80.83	80.63
J48	68.39	66.84	71.50	68.91	56.48	72.53	73.06	66.84	68.07
Kstar	76.68	83.41	82.38	83.42	60.62	82.38	80.83	79.79	78.69
Logistic	83.41	76.16	75.67	76.16	76.68	81.87	86.01	79.79	79.47
LogitBoost	77.20	82.90	81.34	82.90	70.98	76.17	81.35	76.69	78.69
MLP	86	87	88.60	87	82.90	82.90	84.47	83.94	85.35
Naïve Bayesian	78.24	80.82	79.79	80.82	75.65	75.13	77.72	79.79	78.50
Nbtree	63.73	74.61	70.47	74.61	48.70	77.72	75.13	69.43	69.3
PART	66.84	73.10	70.47	71.50	58	68.91	68.39	67.87	68.14
RBF	76.68	80.83	78.24	80.82	68.40	78.24	80.31	78.76	77.79
SMO	84.97	86	82.38	86	83.42	87.05	86.01	84.46	85.04

Table 3. The classifier's accuracies

Another significant observation that can be made is that the addition of the feature selection step has significantly improved the performance accuracy of the AIRS accuracy, an immune-based system. An improvement of almost

7. REFERENCES

[1] R. Dannenberg, J. Foote, G. Tzanetakis, and C. Weare, "Panel: New Directions in Music Information Retrieval," *International Computer Music Conference,*

International Computer Music Association, pp. 52-59, 2001.

[2] G. Tzanetakis and P. Cook, “Musical Genre Classification of Audio Signals,” IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5. 2002.

[3] E. Wold, T. Blum, D. Keislar, and J. Wheaton, “Content-based Classification, Search, and Retrieval of Audio,” IEEE Multimedia, vol.3, no.3, pp. 27-36, 1996.

[4] M. G. Nasuruddin, The Malay Traditional Music. Kuala Lumpur: Dewan Bahasa dan Pustaka, 1992.

[5] N. Norowi, S. Doraisamy and R.W. Rahmat, *Journal of Information Technology in Asia, JITA,* Vol. 2, 2007.

[6] Y. Saeys, I. Inza, and P. Larranaga, A Review of Feature Selection Techniques in Bioinformatics, Bioinformatics Advance Access, and Oxford University press, pgs, Aug. 2007.

[7] K. West, and S. Cox, Features and Classifiers for the Automatic Classification of Musical Audio Signals, In Proceedings of the International *Conference on Music Information Retrieval, ISMIR 2006*.

[8] J. Aucouturier and F. Pachet, “Representing Musical Genre: A State of the Art”', *Journal of New Music Research*. vol. 32, no. 1, pp. 83-93, 1993.

[9] A. Wieczorkowska, “Towards Musical Data Classification via Wavelet Analysis,” *In Proceedings of 12th International Symposium on Methodologies for Intelligent Systems*, pp. 292-300, 2000.

[10] K. Kosina, *Music Genre Recognition*. Diplome Thesis, Hagenberg University, 2002.

[11] S. Dixon, “A Beat Tracking System for Audio Analysis,” *In Proceedings of the Diderot Forum on Mathematics and Music*, Austrian Computer Society, pp. 101-110, 1999.

[12] T. Li and G. Tzanetakis, “Factors in Automatic Musical Genre Classification of Audio Signals,” *IEEE Workshop on Applications of Signal Processing and Audio Acoustics*, WASPAA 2003, pp. 143-146, 2003.

[13] H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, San Francisco, 2nd Edition, 2005.

[14] I. Kaminskyj, “Multi-feature Musical Instrument Sound Classifier,” *Mikropolyphonie, WWW Journal*, Melbourne, Australia, Vol.6, 2001.

[15] S. Lippens, J. Martens, T. Mulder, and G. Tzanetakis, “A Comparison of Human and Automatic Musical Genre Classification,” *In Proceedings of IEEE International Conference on AcousticSpeech and Signal Processing*, Montreal, Canada, 2004.

[16] D. Perrot and R. Gjerdigen, “Scanning the Dial: An Exploration of Factors in Identification of Musical Styles,” In *Proceedings of Society for Music Perception and Cognition*. pp. 88, 1999.

[17] P. Matusky, *Malaysian Shadow Play and Music: Continuity of an Oral Tradition.* Kuala Lumpur: Oxford University Press, 1993.

[18] M. Ang, *A Layered Architectural Model for Music: Malaysian Music on the World Wide Web.* Ph.D. dissertation: UPM, 1998.

[19] J. Becker, “The Percussive Patterns in the Music of Mainland Southeast Asia,” *Ethnomusicology*, vol. 2, no. 2, pp. 173-191, 1968.

[20] N.M. Norowi, S. Doraisamy, and R. Wirza, “ Factors Affecting Automatic Genre Classification: an Investigation Incorporating Non-Western Musical Forms”, *In Proceeding of the 6th International Conference on Music Information retrieval*, pp. 13-21, 2005.

[21] M.A. Hall, and L.A. Smith, “Practical feature subset selection for machine learning”, In Proceedings of the 21st Australian Computer Science Conference, pp. 181–191, 1998.

[22] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” Machine Learning, Vol.46, pp. 389–422, 2002.

[23] D.E. Goldberg, Genetic algorithms in search, optimization and machine learning, Addison-Wesley, 1989.

[24] R. Fiebrink and I. Fujinaga, Feature Selection Pitfalls and Music Classification, *In Proc. Of the International Conference on Music Information Retrieval*, *ISMIR 2006*, pgs 340-341, 2006.

[25] A. Watkins, J. Timmis, and L. Boggess. Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm. *Genetic Programming and Evolvable Machines*, 5(3):291-317, 2004.

[26] J. Brownlee. WEKA Classification Algorithms, Version 1.6, http://sourceforge.net/projects/wekaclassalgos., Accessed December 2007.

[27] E. Frank and I. H. Witten. Generating Accurate Rule Sets without Global Optimization. *In Fifteenth International Conference on Machine Learning*, Morgan Kaufmann, 1998.

[28] S. S. Keerthi, et al. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation,* 13(3):637-649, 2001.

[29] De Castro, L.N. Timmis, J, Articial Immune System as a novel soft computing paradigm, *Soft Computing Journal 7 (7)*, 2003.

[30] Watkins, A., AIRS: A Resource Limited Artifiial Immune Classifier, *M.S. thesis, Dept. of Comp. Sc, Mississippi State University,* 2001.