

# A Class Centric Feature and Classifier Ensemble Selection Approach for Music Genre Classification

Hasitha Bimsara Ariyaratne, Dengsheng Zhang, and Guojun Lu

Monash University, Gippsland School of IT, Churchill, Australia  
{hasitha.ariyaratne,dengsheng.zhang,guojun.lu}@monash.edu

**Abstract.** Music genre classification has attracted a lot of research interest due to the rapid growth of digital music. Despite the availability of a vast number of audio features and classification techniques, genre classification still remains a challenging task. In this work we propose a class centric feature and classifier ensemble selection method which deviates from the conventional practice of employing a single, or an ensemble of classifiers trained with a selected set of audio features. We adopt a binary decomposition technique to divide the multiclass problem into a set of binary problems which are then treated in a class specific manner. This differs from the traditional techniques which operate on the naive assumption that a specific set of features and/or classifiers can perform equally well in identifying all the classes. Experimental results obtained on a popular genre dataset and a newly created dataset suggest significant improvements over traditional techniques.

**Keywords:** music retrieval, feature selection, classifier ensemble, music genre classification.

## 1 Introduction

Recent advancements in digital media encoding, storing and delivering technologies have led to a significant increase in the number of digital audio files. As a result, managing music has become a challenging task. Genre is the most widely used descriptor in organizing and searching large music collections [1]. It has been shown that existing audio features and classifiers have reached a “glass ceiling” [2], because most new features and classifiers only show a marginal improvement. Further investigation of existing literature reveals that majority of the methods use a single classifier trained with a collection of different audio features. Some of the recent works have investigated the effectiveness of employing attribute selection methods for feature combination [3], while others [4,5] have demonstrated the effectiveness of using classifier ensembles to improve classification accuracy. However, these techniques operate on the assumption that a specific set of features and/or a specific set of classifiers can perform well for all music classes. But studies [5] have shown class-specific feature selection can produce better performance. In this paper we take the existing research one

step further by proposing a class centric feature and classifier ensemble selection method.

Music genre classification has two steps: audio feature extraction and classification. Audio features are usually designed only to capture specific qualities of sound; therefore, they are limited in terms of generalization. For example, features like Zero Crossing Rate (ZCR), and Linear Predictive coefficients (LPC) can be used to discriminate instrumental music (e.g. Classical) from vocal music (e.g. Country) due to their ability in capturing certain characteristics of human voice. However, they perform poorly at discriminating between pure instrumental genres (e.g. classical or jazz). The most common way to solve this problem is to combine different features. However, fusing features should be done carefully, since not all features contribute to the classification task equally.

Moreover, large feature vectors can also lead to high computational complexities and complications (generally known as the “curse of dimensionality”). Therefore feature selection based on individual classes can provide an effective way to reduce the number of features while preserving the discrimination power. Most traditional techniques rely on multiclass classifiers working on all the classes; therefore, class specific feature selection is not performed. An alternative to this approach is the use of hierarchical classifiers where a multiclass problem can be split into a set of smaller problems which can be optimized individually. Zhang et.al [6] proposed the use of a manually generated classification hierarchy and a set of manually selected features for each node, while Silla et.al [7] proposed a node specific classifier and feature selection technique based on a predefined genre taxonomy, however neither have provided empirical results on the effectiveness of hierarchical classification against traditional techniques. Ariyaratne et.al [8] has proposed a hierarchical classification technique based on an auto generated tree which utilizes node specific feature and classifier selection, their empirical results have shown the hierarchical approach to perform better, however despite high accuracies obtained at higher levels, the leaf level accuracies were not significantly improved due to errors at higher levels propagating downward. Therefore in this paper we focus on a nonhierarchical approach to class centric classification based on a binary decomposition technique.

There are two main techniques for breaking down a multiclass problem into a set of binary classification problems: One-Vs-One (OVO) and One-Vs-All (OVA). OVA is the most simplest of the two, this technique creates  $n$  binary classifiers for each  $n$  classes where one class is considered positive while all other classes are considered negative. In comparison OVO technique builds  $n(n-1)/2$  classifiers for each possible combination of class pairs. These two approaches adopt different techniques to aggregate the results of multiple classification problems to decide the final outcome [9].

The rest of the paper is organized as follows: Section 2 presents the proposed method which utilizes a binary decomposition technique for disassembling the multiclass problem in order to perform class centric feature and classifier selection. Section 3 presents experimental results and the paper is concluded in section 4.

## 2 The Proposed Method

In this section we present an overview of our proposed class centric feature and classifier ensemble selection method. As indicated earlier, in order to perform class centric feature and classifier selection, we need to address each class separately. Both OVA and OVO binary decomposition techniques let us focus on each class individually. Galar et.al [9] has compared the performance of several different classifiers for solving multiclass problems using OVA and OVO decomposition methods on number of publicly available nonmusical datasets. They have concluded that OVO methods generally perform better. Similar evidence can be seen in Silla et.als work [10] where a new feature selection technique was introduced for music genre classification. They have evaluated the performance on a set of classifiers under OVA and OVO schemes. Both previous work have used a relatively low number of features (less than 40), therefore we conducted our own experiments to validate these claims with large number of features(183 features). We adopted a voting aggregation method for OVO and maximum confidence level for OVA. Experimental results are shown in Table 3 under section 3.1. As indicated in previous work and as confirmed by our own experiments, OVO noticeably outperforms OVA, hence we adopted the one-vs-one decomposition technique for our class centric approach. Fig. 1 shows an overview of our proposed method. We start by disassembling the multiclass genre classification problem into a set of binary problems. Training data sets for each problem are constructed of instances belonging to the corresponding class pairs. These datasets are then preprocessed by performing feature selection to reduce dimensionality and improve predictive accuracy (by removing irrelevant features that can introduce noise into the data).

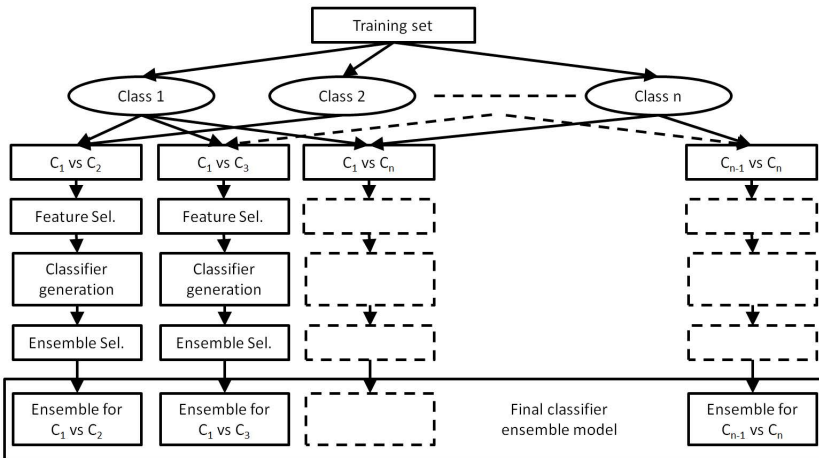


Fig. 1. Overview of the proposed method

Afterwards, we generate a set of classifiers for each problem and then choose the best classifier ensemble for each task through a classifier ensemble selection method outlined in following section.

## 2.1 Classifier Ensemble Approach

There are many classifier ensemble techniques in the literature such as boosting, bagging, random forests and stacking [11]. The main idea behind the use of classifier ensembles is to combine multiple classifiers to improve classification accuracy. However, as with feature selection, we hypothesize that selecting the best set of classifiers out of a multitude of classifiers for each classification problem is more important than randomly combining a set of classifiers in a multiclass setting. Therefore we propose the use of a classifier ensemble selection technique to choose the best performing set of classifiers (classifier ensemble) for each classification task. Caruana et.al [12] have proposed a classifier ensemble selection technique to select a set of classifiers from a multitude of model libraries. We have adopted this approach to choose the best classifier ensemble. The basic steps of the algorithm are as follows:

- Separate training instances into individual classes
- Construct  $n(n-1)/2$  datasets for OVO comparisons with each pair of classes
- Carry out feature selection for each dataset
- Generate a collection of classifiers with bagging and selection with replacement
- Perform feature selection for each dataset based on the classifier optimum feature selection method.
- Train the classifiers on the classifier optimal set of features
- Select the best classifier ensemble for each binary class by:
  - Add to the ensemble, the classifier model which maximizes the ensembles' overall performance (measured using a certain error metric: i.e accuracy, RMS etc.) in a hill climbing fashion. The overall performance is computed by averaging the performance of each classifier.
  - Repeat the above step until all the models have been examined.
  - Return the ensemble that has maximum performance.

## 3 Experimental Results

In this section we present the details of the experiments carried out to measure the impact of class-centric feature and classifier selection. As mentioned earlier, previous work in MIR research has produced a wide variety of features, different authors use different sets of audio features, however there are no "correct" set of features for any particular MIR task. In our work, we chose a set of 13 widely used features which are robust and computationally efficient low level features as listed in Table 1. They comprise of six MPEG7 features and seven other most commonly used audio features. We derive 4 statistical properties: i.e. mean,

**Table 1.** Audio features used for the experiments

Feature Name	No.of features
1. Total Energy	4
2. Fundamental Frequency	4
3. Loudness Sensation	32
4. Integral Loudness	4
5. Audio Spectral Centroid	4
6. Spectral Rolloff	4
7. Audio Spectrum Spread	4
8. Audio Spectrum Flatness	16
9. Audio auto correlation	13
10. Log Attack Time	1
11. Temporal Centroid	1
12. Zero Crossing Rate	4
13. Mel Frequency Cepstral Coefficients	96
Total	183

**Table 2.** Details of the HBA and GTZAN datasets

	HBADS	GTZAN
No. of genres	15	10
Genres	Blues, Classical, Country, Disco, Hiphop, Indian, Jazz, Metal, Opera, Pop, Reggae, Rock, Salsa, Techno, Ambient	Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock
No. of files	7500 (500 per genre)	1000 (100 per genre)
Format	22,000KHz, 64kbps MP3	22,000KHz, 16bit PCM
Size	2.6GB	1.2GB

variance, covariance, and their numerical partial derivatives (i.e. the differences between successive elements).

Features 1, 10, 11 and 12 are time domain features while the rest are extracted in the frequency domain. MFCC is a widely used feature in many different areas of MIR, for our experiments we extracted the first 24 coefficients and their previously mentioned statistical properties. Features 2 ,5 through 8 provide various statistical measurements related to the frequency spectrum of a sound. Features 3 and 4 capture the human perception of loudness while auto correlation feature can be used to analyze reoccurring patterns (i.e beats, tempo) in a signal. We used two music datasets in our experiments. As our first dataset we chose the widely used genre benchmarking dataset GTZAN[13], it contains 10 genres with 100 clips each. In order to test the robustness of the proposed method on a larger dataset, and due to the lack sufficiently large datasets in the literature, we constructed our own dataset by extending the genre set of GTZAN to 15. The new dataset contains a carefully selected set of songs with no duplicates or ambiguous class labels. The genre labels are taken directly from official album/song information to ensure maximum accuracy of the labels. Details of both datasets are given in Table 2.

### 3.1 Experimental Setup

We used the WEKA[14] machine learning platform as the test bed for conducting our experiments. We chose the following 5 well known classifiers in our ensemble approach: K-Nearest Neighbour (IBK), Naive Bayes (NB), Decision Tree (J48), Logistic Regression (LR), and Sequential Minimal Optimization: Support Vector Machine (SMO).

As mentioned in section 2 we conducted our own experiments to choose the best binary multiclass decomposition scheme between OVA and OVO with the GTZAN dataset. Results are presented in table 3.

**Table 3.** Classifier performance between OVA and OVO

Classifier	One vs All	One vs One
IBK	54.70	<b>57.64</b>
NB	47.35	<b>56.74</b>
J48	48.82	<b>55.94</b>
SMO	50.52	<b>67.35</b>

From these results we can conclude that OVO scheme has a notable advantage over OVA. Therefore we primarily focus on OVO for implementing our proposed approach. There are many feature selection algorithms available in the literature which can be mainly classified into two groups: Filter methods and Wrapper methods [14]. Filter methods rely on general characteristics of the data; they attempt to measure the importance of each feature or feature subset using a score metric such as information gain, Chi-squared distributions or correlation coefficients, then choose (“Filter”) the best set. In contrast, wrapper methods make use of a learning algorithm to search the feature space and evaluate the usefulness of features for the classification problem at hand, in other words the learning algorithm is “Wrapped” into the feature selection process. The later produce a more optimum set of features; however, they are much more computationally expensive and prone to over fitting. Sequential forward/backward selection and hill climbing methods are some of the search methods commonly used for searching the feature subsets in wrapper methods [14]. Since our primary aim is to investigate the effectiveness of class centric feature and classifier selection, we choose the more computationally efficient filter based approach. We used the following 4 widely used feature selection (FS) techniques [3]: Correlation based Feature Selection (CSF), Chi-square Feature Evaluation (CHI), Gain Ratio Feature Evaluation (GAIN), and Principal Component Analysis (PCA).

Table 4 list overall accuracies obtained for each classifier with and without feature selection on GTZAN dataset. This experiment was conducted to analyze the impact of different feature selection techniques on different classifiers, and choose the best technique for each classifier.

As seen in table 4, feature selection methods do have a significant impact on majority of the learning algorithms. These results also agree with the findings of [3]. One interesting observation is how PCA has negatively impacted the accuracy. Even though PCA can be used to reduce the dimensionality of data;

**Table 4.** Impact of feature selection on classifiers - HBA dataset

Classifier	No FS	CSF	CHI	GAIN	PCA
IBk	53.05	<b>59.60</b>	56.66	56.66	31.76
NB	58.94	<b>62.74</b>	57.64	57.64	54.31
J48	50.45	51.17	51.37	<b>51.37</b>	31.56
LR	65.00	<b>69.45</b>	68.63	68.12	67.56
DT	<b>41.05</b>	31.01	35.88	35.88	32.15
SMO	69.56	66.07	70.00	<b>70.00</b>	60.98

strictly speaking, PCA is not a feature selection technique, rather a feature extraction method that constructs new features through linear projection of high dimensional vector into a low dimensional vector while retaining characteristics of the dataset that contribute to most variation. Therefore even though PCA extracted features are optimal for reconstructing original data, they are not always better for classification [15]. We used these results as a guide for choosing the best feature selection technique for different classifiers in our ensemble approach.

### 3.2 Classifier Ensemble Approach

We adopt the classifier ensemble selection technique proposed in [[12]] under one-vs-one binary decomposition strategy to improve classification accuracy by narrowing down the selection criteria focusing on each class. Once the final classifier ensemble model has been constructed, classification for an unlabelled test instance is performed by choosing the class which generated majority votes among all  $n(n-1)/2$  OVO problems. For each OVO problem, classifier ensembles are also evaluated based on majority voting. Results obtained for both datasets are listed in tables 5 and 6. The columns are labelled as follows:

- NoFS: Lists classification accuracies for each classifier on its own (without any feature selection).
- FS: Best accuracy obtained with feature selection for each classifier (not class centric).
- CCFS: Accuracies obtained for class centric feature selection through OVO decomposition
- CES: Accuracy of classifier ensemble selection (i.e conventional multiclass problem with classifier ensemble selection),
- CCFS+CES: Accuracy of class centric feature selection and classifier ensemble selection based on OVO decomposition (our proposed approach).

From these results we can observe that classification accuracies are improved across majority of the classifiers when class specific feature selection is performed, except for support vector machine. The implementation of SMO Support Vector Machine already performs OVO binary comparisons when handling multiclass problems. Therefore using a SMO classifier in an OVO setting has no extra advantage. Furthermore, a support vector machine is already a very efficient

**Table 5.** Comparison of classification accuracies between different classification techniques on the HBA Dataset

Classifier	No FS	FS	CCFS	CES	CCFS+CES
IBk	53.05	59.60	62.74		
NB	58.94	62.74	63.20		
J48	50.45	51.37	59.11	71.37	<b>77.88</b>
LR	65.00	69.45	71.30		
DT	41.05	35.88	38.80		
SMO	69.56	71.00	69.70	70.11	76.56

**Table 6.** Comparison of classification accuracies between different classification techniques on the GTZAN Dataset

Classifier	No FS	FS	CCFS	CES	CCFS+CES
IBk	54.70	64.41	66.70		
NB	57.05	60.88	61.80		
J48	49.41	49.70	57.11	65.60	<b>70.88</b>
LR	50.0	61.05	64.00		
DT	41.47	40.20	42.61		
SMO	66.17	72.05	69.70	70.11	71.0

and highly optimized multi class classifier; this is why adding SMO into the classifier ensemble didn't have a significant impact on the ensemble approach either.

Further analysis shows that classifier ensemble selection(CES) performs slightly higher (71.30% and 65.60%) than the best individual classifier: CF (69.45% and 64.41%) for HBA and GTZAN datasets respectively. Finally we can see a significant improvement when class specific ensemble classifier selection is performed. The improvement is about 6.5% for the HBA dataset and 5.28% for the GTZAN dataset. From these results we can conclude that the use of a class centric feature selection and classifier ensemble selection of otherwise weak classifiers can perform equally or better than a highly efficient classifier such as the support vector machine.

## 4 Conclusion

In this paper we proposed a class centric feature and classifier ensemble selection technique for music genre classification. We presented experiments to validate the selection the best multiclass decomposition technique and the best feature selection technique for music genre classification. Building upon these findings we tested our proposed method using two genre datasets and a set of weak classifiers and low level features. The promising results obtained through experiments validated our initial hypothesis that a class centric feature selection combined with a classifier ensemble selection can improve genre classification accuracy. In this work we tested our hypothesis using the most commonly used feature



selection techniques (i.e: filter methods), aggregation of binary decomposition methods (i.e: majority voting) and classifier ensemble selection method (hill climbing forward selection by optimizing for accuracy). However better alternative techniques do exist and they may further improve performance, therefore need further investigation.

## References

1. McKay, C., Fujinaga, I.: Musical Genre Classification: Is it worth pursuing and how can it be improved? In: Proc. ISMIR (2006)
2. Aucouturier, J., Pachet, F.: Improving timbre similarity: How high is the sky? *Journal of Negative Results in Speech and Audio Sciences* 1(1) (2004)
3. Doraisamy, S., Golzari, S., Norowi, N.M., Sulaiman, M.N., Udzir, N.I.: A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music. In: Proc. ISMIR (2008)
4. Yaslan, Y., Cataltepe, Z.: Audio Music Genre Classification Using Different Classifiers and Feature Selection Methods. In: Proc. ICPR (2006)
5. Soares, C., Williams, P., Gilbert, J.E., Dozier, G.V.: A Class-Specific Ensemble Feature Selection Approach for Classification Problems. In: Proc. ACMse (2010)
6. Zhang, T.: Semiautomatic approach for music classification. In: Proc. SPIE Conf. on Internet Multimedia Management Systems (2003)
7. Silla, C., Freitas, A.: Novel top-down approaches for hierarchical classification and their application to automatic music genre classification. In: Proc. IEEE SMC (2009)
8. Ariyaratne, H.B., Zhang, D.: A novel automatic hierarchical approach to music genre classification. In: Proc. ICME Workshop on Advances in Large Scale Multimedia Data Collection, Mining and Retrieval (2012)
9. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F.: An Overview of Ensemble Methods for Binary Classifiers in Multiclass Problems: Experimental Study on One-vs-One and One-vs-All Schemes. In: *Pattern Recognition*, vol. 44(8), pp. 1761–1776. Elsevier Science Inc. (2011)
10. Silla Jr., C.N., Koerich, A.L., Kaestner, C.A.A.: Feature Selection in Automatic Music Genre Classification. In: Proc. ISM (2008)
11. Rokach, L.: Taxonomy for Characterizing Ensemble Methods in classification Tasks: a review and annotated bibliography. In: Azen, S.P., Kontoghiorghes, E.J., Lee, L.C. (eds.) *Computational Statistics & Data Analysis*, vol. 53(12), pp. 4046–4072 (2009)
12. Caruana, R., Niculescu-Mizil, A., Crew, G., Ksikes, A.: Ensemble selection from libraries of models. In: Proc. ICML (2004)
13. Tzanetakis, G., Cook, P.: Musical Genre Classification of Audio Signals. *IEEE Speech Audio Process* 10(5), 293–302 (2002)
14. Witten, I., Frank, E., Hall, M.: *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd edn. Morgan Kaufmann (2011)
15. Cevikalp, H.: *Feature Extraction Techniques in High-dimensional Spaces: Linear and Nonlinear Approaches*. Ph.D Thesis, Vanderbilt University (2005)