# AN ASSOCIATION-BASED APPROACH TO GENRE CLASSIFICATION IN MUSIC

**Tom Arjannikov**
University of Lethbridge
tom.arjannikov@uleth.ca

**John Z. Zhang**
University of Lethbridge
zhang@cs.uleth.ca

## ABSTRACT

Music Information Retrieval (MIR) is a multi-disciplinary research area that aims to automate the access to large-volume music data, including browsing, retrieval, storage, etc. The work that we present in this paper tackles a non-trivial problem in the field, namely music genre classification, which is one of the core tasks in MIR. In our proposed approach, we make use of association analysis to study and predict music genres based on the acoustic features extracted directly from music. In essence, we build an associative classifier, which finds inherent associations between content-based features and individual genres and then uses them to predict the genre(s) of a new music piece. We demonstrate the feasibility of our approach through a series of experiments using two publicly available music datasets. One of them is the largest available in MIR and contains real world data, while the other has been widely used and provides a good benchmarking basis. We show the effectiveness of our approach and discuss various related issues. In addition, due to its associative nature, our classifier can assign multiple genres to a single music piece; hopefully this would offer insights into the prevalent multi-label situation in genre classification.

## 1. INTRODUCTION

The recent advances in technology, such as data storage and compression, data processing, information retrieval, and artificial intelligence, facilitate music recognition, music composition, music archiving, etc. The Internet is further promoting the enormous growth of digital music collections. Millions of songs previously in physical formats are now readily available through instant access, stimulating and motivating research efforts in meeting new challenges. Among them is *Music Information Retrieval (MIR)*, an interdisciplinary area that attracts practitioners from information retrieval, computer science, musicology, psychology, etc. One of the main tasks in MIR is the design and implementation of algorithmic approaches to managing large collections of digital music, including automatic tag annotation, recommendation, playlist generation, etc.

The work to be presented in this paper explores the feasibility of applying association analysis to music genre classification. Through our experience with music data, we have found that there are some inherent associations between audio characteristics and human assigned music genre labels. Accordingly, it would be desirable to see whether these associations, if found, can provide insight into genre classification of music. Our work in this paper is geared toward this target.

In a nutshell, our proposed approach uses music data itself by extracting useful information from it and conducting association analysis to make genre prediction. When we talk about the actual sound data of music, we refer to whatever is stored on various media, such as magnetic tapes and now in the digital format. We can extract useful information from this data via signal processing. This information represents the different characteristics of the actual sound stored on media [10]. We refer to it as *content-based features* and use it with our approach. To our knowledge, we are among the first to propose using association analysis for music genre classification in the MIR community.

## 2. PREVIOUS WORK

### 2.1 Classification in MIR

Classification is the process of organizing objects into predefined classes. It is a supervised type of learning, where we are given some labeled objects from which we form a computational model that can be used to classify new, previously unseen objects [15].

Classification is one of the core tasks in MIR, since it is usually the first step in many applications, such as on-line music retrieval, playlist recommendation, etc. In our work, we focus on genre classification, which is concerned with categorizing music audio into different genres. Tzanetakis and Cook [18] are among the first to work on this problem, where the task is to label an unknown piece of music with a correct genre name. They show that this is a difficult problem even for humans and report that college students achieve no more than 70% accuracy.

Previous works in MIR along this direction include the following. DeCoro *et al.* [5] use *Bayesian Model* to aid in hierarchical classification of music by aggregating the results of multiple independent classifiers and, thus, perform error correction and improve overall classification accuracy. Recent examples of using *Support Vector Machines (SVM)* for music genre classification include an investigation of Meng and Shawe-Taylor [13], where they explore different kernels used in a support vector classifier. Li and Sleep [9] extend normalized information distance into kernel distance for SVM and demonstrate classification accuracy comparable to others. In addition, recently, Anglade *et al.* [3] use *Decision Tree* for music genre classification by utilizing frequent chord sequences to induce context free definite clause grammars of music genres.

### 2.2 Association Analysis in MIR

Association analysis attempts to discover the inherent relations among data objects in an application domain. These relations are represented as association rules. An example of such application domain is the shopping basket analysis in supermarkets, where one tries to discover relations among the items purchased by customers. For example, the association rule $\{milk, eggs\} \rightarrow \{bread\}$ implies that, if *milk* and *eggs* are bought together by a customer, then *bread* is likely to be bought as well, i.e., they have some inherent statistical relationships [7].

We consider the so-called itemsets, such as $\{milk, eggs, bread\}$ in the above example, to be frequent if they appear in many transactions. The *support* of an itemset represents the percentage of transactions that contain the itemset and *minimum support* is the threshold that separates the frequent itemsets from the infrequent ones. A frequent itemset can produce an association rule of the form $A \rightarrow B$, where $A$ and $B$ are non-empty itemsets and $A \bigcap B = \phi$. An association rule holds for a dataset with some minimum support and *confidence*, which is the percentage of transactions containing $A$ that also contain $B$ [7].

A formal treatment of applying association analysis in MIR is in Section 3. Within the context of MIR, each track or music piece is represented using a set of content-based features derived from its digitized data. Together, a set of these features place the given track in a discrete location in the feature space. Intuitively, the tracks that are very similar to each other may share the same neighborhood. This could help with organizing music collections for effective data retrieval. When grouped together, the features contain some patterns. We would like to look for these patterns and use them for music genre classification.

Kuo *et al.* [8] propose a way to recommend music based on the emotion that it conveys and look for associations in data that contains information perceived only by humans. Similarly, Xiao *et al.* [19] use a parameterized statistical model to look for associations between timbre and perceived tempo. Liao *et al.* [12] use a dual-wing harmonium model to discover association patterns between MTV video clips and the music that accompanies those clips. Neubarth *et al.* [14] present a method of association rule mining with constraints and discover rules in the form of $A \rightarrow B$, telling that either region implies genre or genre implies region. Arjannikov *et al.* [4] use association analysis to verify tag annotation in music, though their approach is based on textual music tags and is not content-based. Our work to be presented below is different from the above and is among the initial efforts to apply association analysis to content-based music genre classification.

## 3. CLASSIFYING MUSIC INTO GENRES VIA ASSOCIATION ANALYSIS

Our work in this paper is focused on the music genre tags. As stated in [6, 10, 10], any discrete set of tags that are not correlated can be used as categories, or classes, into which we could split a collection of music pieces. Arjannikov *et al.* [4] show that association analysis reveals patterns in music textual tags. This motivates our investigation of association analysis in content-based music features.

### 3.1 Notation

Association analysis requires discrete items, however, most content-based music features are not. Thus, when given a set of features $F = \{f_1, f_2, f_3, \cdots, f_k\}$, we discretize each feature into a predetermined number of bins $b$, where $b > 1$, and derive a new feature set $F' = \{f'_{1_1}, f'_{1_2}, \cdots, f'_{1_b}, f'_{2_1}, f'_{2_2}, \cdots, f'_{2_b}, \cdots, f'_{k_1}, f'_{k_2}, \cdots, f'_{k_b}\}$. Then, from the set of music pieces $M$, we derive a transactional style dataset $D = \{d_1, d_2, \cdots, d_r\}$, where $r = |M|$. Each transaction $d_i = \{a_1, a_2, \cdots, a_k\}$ corresponds to a music piece and each $a_j$ in $d_i$ is a feature item in the literal form $F_pB_q$, where $p$ corresponds to the feature number in $F'$ and $q$ corresponds to the bin number, into which the feature for the particular music piece falls. For example, if the first content-based feature is a number between 0 and 1, and it is discretized into 10 equidistant bins, then, given a particular music piece, whose first feature value is $0.125$, its corresponding $d_i$ would contain the label $F_1B_2$.

When we formulate our problem as described above, the music set $M$, becomes a transactional set $D$ suitable for association mining.

### 3.2 Proposed Approach

We call our proposed approach *association-based music genre classifier (AMGC)*. Figure 1 depicts the whole process of using AMGC, which is detailed below.

#### *3.2.1 AMGC*

We start by preparing our data during the pre-processing stage. First, we acquire content-based features from music; in this paper, we use the features that have already been extracted and published for the purpose of comparing different classifiers on even ground [16, 17]. Then, we discretize any continuous features. It is worth noting that obtaining optimal discretization is an open problem in machine learning. In our work, we use feature discretization based on equal width of bins, for its simplicity, to avoid any possible bias based on class labels. Then we form transactional style datasets, as described in Section 3.1, and split the training dataset into subsets, one for each genre. Finally, we remove any items that appear in all transactions with a certain *frequency threshold* (FRQ), which is the percent of transactions containing the item.

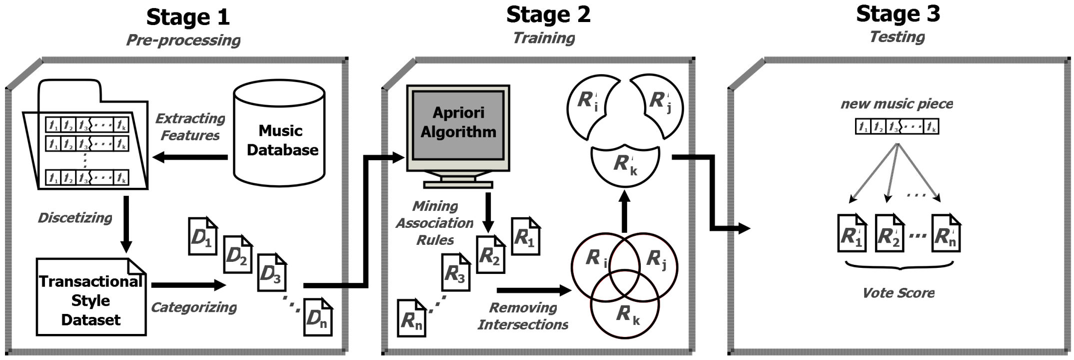


**Figure 1**. The three stages of our proposed association-based approach to classify music into genres.

During the training stage, we invoke the *Apriori* algorithm [1, 2] and mine frequent itemsets from each genre's sets of items at some minimum support. From these we generate classification rules of the form $A \rightarrow B$, where $A$ is the frequent itemset and $B$ is the genre associated with that itemset. Then, we remove any itemsets that appear in two or more genres. The resulting rules uniquely represent their respective genres and we use them for classification during the last stage.

#### 3.2.2 Scoring Method

To obtain a classification score for each genre, we use the following four components. *Itemset Percentage* (IP) is the percent of itemsets that a given music piece matches for a given genre out of all itemsets matched from that genre. *Support Sum* (SS) is the sum of the matched itemsets' minimum support divided by the sum of all itemsets' minimum support for the given genre. *Confidence Sum* (CS) is the current genre's confidence sum of the matched itemsets divided by the sum of all itemsets' confidence. Finally, *Length Sum* (LS), the sum of cardinalities of the matched itemsets divided by the sum of cardinalities of all itemsets for the given genre.

We score each music piece against each genre's set of rules as following. First, we create a voting vector, whose cardinality is equal to the number of genres, and compute the corresponding component's value for each genre. Then, the genre with the highest value is voted as a candidate of that component, and its element in the voting vector is incremented by 1. Thus, the four components result in four votes and the genre with the highest number of votes is declared as winner and becomes the predicted genre of the given music piece.

#### 3.2.3 Accuracy Evaluations

In our work, we use the following classification measures. *Recall*, also known as *sensitivity*, represents the percentage of correctly classified instances for that genre [7]. *Precision* reflects the percentage of correctly classified instances from all instances that are perceived as belonging to that genre by the classifier [7]. Finally, *accuracy* is calculated by dividing the number of all correctly classified instances for all genres by the total number of predictions made [7].

Because AMGC can assign multiple genre labels to a single music piece, we compute the *Multi-Labeling Rate* (MLR) by dividing the total number of predicted labels by the number of all test instances of a genre. MLR falls into the range between $1$ and the total number of genres with frequent itemsets. The closer it is to 1, the fewer multi-label assignments were made, which indicates that AMGC is performing more like a single-label classifier. If MLR is equal to the total number of genres, then the results of classification are least useful. Furthermore, if MLR is below $1$, then there are music pieces, whose genres could not be predicted.

### 3.3 Goals

Our aim is to test whether the classification rules obtained from music content-based features by AMGC can be used to categorize music into genres. For this, we designate three goals: ($G_1$) AMGC achieves a classification accuracy that is better than choosing genres at random; ($G_2$) AMGC is stable - when given similar datasets, it should achieve similar classification accuracy; ($G_3$) AMGC attains higher accuracy with better quality data and fewer genres.

## 4. EXPERIMENT RESULTS AND DISCUSSIONS

### 4.1 Data Preparation

The classification task at hand requires content-based features paired with genre tags and we find two datasets that fit this requirement.

The *Latin Music Database* [17], denoted as $D_{\mathrm{LMD}}$, is popular in the music genre classification task despite its small size. There are many classification results available in the literature, which are based on a set of features that has already been extracted and circulated as part of $D_{\mathrm{LMD}}$. Thus, we can test the feasibility of our approach without introducing variance based on difference in feature extraction techniques. Moreover, $D_{\mathrm{LMD}}$ usually results in high classification accuracy for many methods [17]. We use one of the three sets of features included with it, which is extracted from the beginning 30 seconds of each music piece.

The *Million Song Dataset Benchmarking* [16], denoted as $D_{\mathrm{MSDB}}$, is much larger than $D_{\mathrm{LMD}}$ and boasts several sets of content-based features. We use five of these sets and the genre labels, which were originally obtained from Allmusic [16]. Additionally, we restrict the number of tracks to 1000 per genre, in order to balance the number of training and testing examples among genres.

| Dataset name | Number of songs | Number of genres | Number of features | Type of Features |
|---|---|---|---|---|
| $D_{\text{LMD}}$ | 3000 | 10 | 30 | MFCC |
| $D_{\text{MSDB-1}}$ | 1500 | 15 | 10 | MM |
| $D_{\text{MSDB-2}}$ | 1500 | 15 | 16 | Spectral |
| $D_{\text{MSDB-3}}$ | 1500 | 15 | 20 | LPC |
| $D_{\text{MSDB-4}}$ | 1500 | 15 | 20 | AM |
| $D_{\text{MSDB-5}}$ | 1500 | 15 | 26 | MFCC |

**Table 1**. Music genre datasets and their statistics.

We include some statistical information about the datasets in Table 1 and label them accordingly. We split each one into two equal-sized partitions at random, while maintaining the genres balanced; each genre is represented by equal number of tracks in both partitions. One of the partitions becomes the training set and the other becomes the testing set. If there are too many music pieces belonging to one genre as compared to others, we remove the extra tracks at random. If a genre is represented by fewer pieces than 300 for $D_{\text{LMD}}$ and 1000 for $D_{\text{MSDB}}$, then we do not use that genre in our experiments. This reduces the original $D_{\text{MSDB}}$ dataset to 17 genres from 25. Moreover, during Stage 2 of our proposed approach, when we mine frequent itemsets, two of the genres produce none; therefore, only 15 genres persist, as reported in Table 1. $D_{\text{LMD}}$ remains at 10 genres because it was originally balanced at 300 music pieces per genre.

In the following section, we demonstrate through our experiment results how we achieve the three goals formulated in Section 3.3.

### 4.2 Results and Discussions

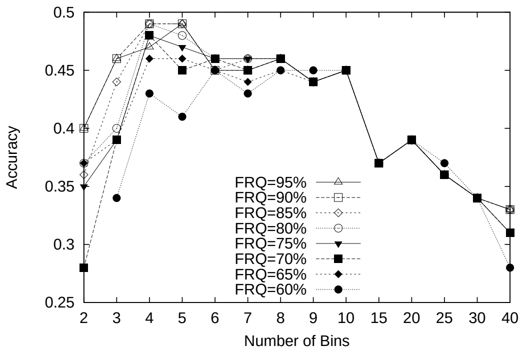


**Figure 2**. $D_{\text{LMD}}$ at minimum support $= 20\%$.

During our experiments, we observe that our proposed parameters affect the classification accuracy, and thus, they are effective. It is evident from Figures 2 and 3 that the number of discretization bins affects the classification accuracy for both $D_{\text{LMD}}$ and $D_{\text{MSDB}}$. Figure 4 demonstrates how the classification accuracy is affected by the minimum support parameter. We also note that AMGC performs

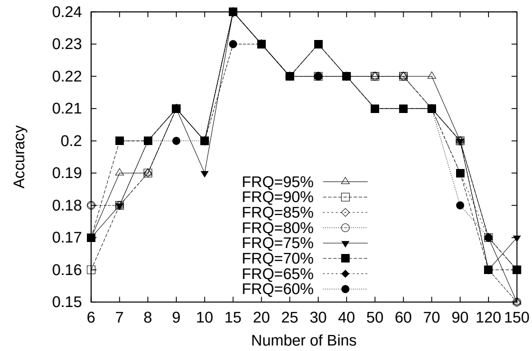


**Figure 3**. $D_{\text{MSDB-2}}$ at minimum support $= 2\%$.

much better than if we were to choose genres at random. Thus, we confirm that AMGC works for some parameter settings and conclude our work towards $G_1$.

As demonstrated in the literature, the classification accuracy usually increases when the number of classes is reduced [11]. Thus, we reduce the number of genres for both $D_{\text{LMD}}$ and $D_{\text{MSDB}}$ to 5 and observe that AMGC performs better. Therefore, we report only the results for the smaller set of genres in Figures 4 through 9. We also observe that $D_{\text{LMD}}$ achieves higher accuracy than $D_{\text{MSDB}}$ as can be seen in Figures 2 and 3. This concludes our work towards $G_3$, as AMGC performs better with a better quality dataset, moreover, it performs better on a reduced set of genres.

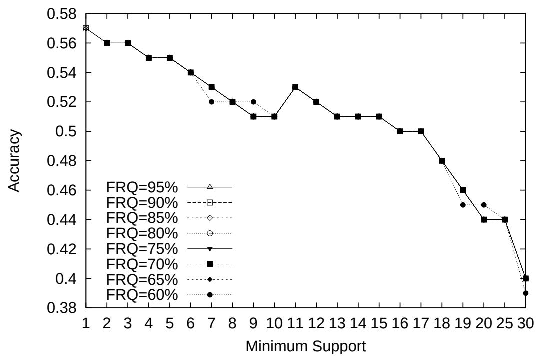


**Figure 4**. $D_{\text{MSDB-2}}$ with $number\ of\ bins = 20$.

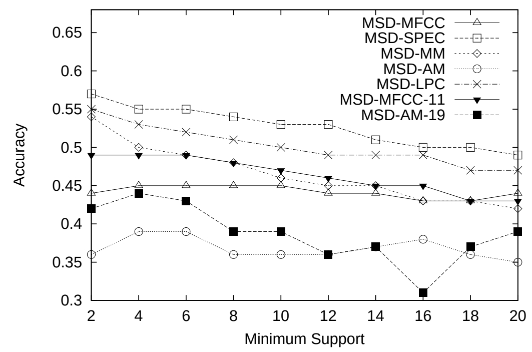


**Figure 5**. All five $D_{\text{MSDB}}$ datasets compared, with $number\ of\ bins = 13$, unless otherwise specified.

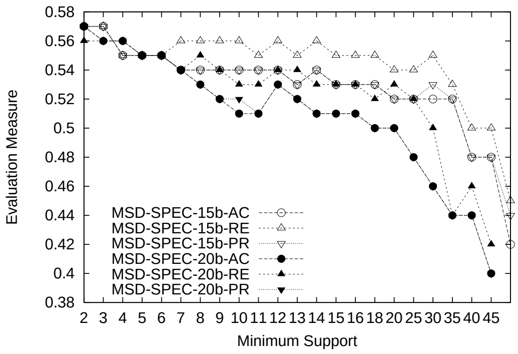


**Figure 6**. $D_{\text{MSDB-2}}$ across different minimum support.

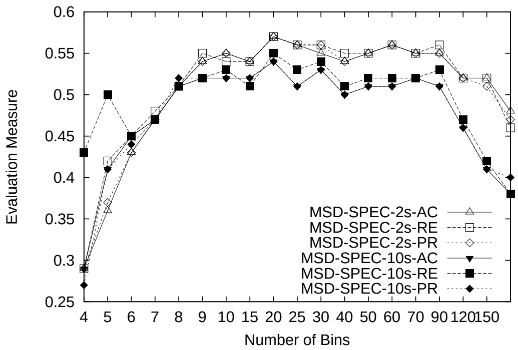


**Figure 7**. $D_{\text{MSDB-2}}$ across different number of bins.

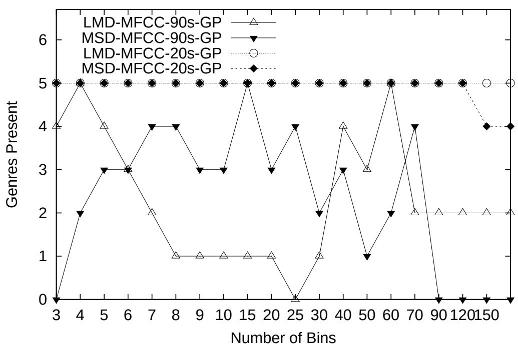


**Figure 8**. 5 genres of $D_{\text{LMD}}$ and $D_{\text{MSDB-5}}$ at $FRQ = 95$.

It is clear from Figures 2, 3 and 4, that the FRQ parameter does not significantly affect the classification accuracy, although, it produces highest accuracy overall when set to 95%. We use this setting in all of the experiment results in Figures 5 through 9.

During our experiments, we observe that $D_{\text{MSDB}}$ datasets perform best at lower minimum support and number of bins settings. We set the number of bins to 13 and perform a sweep across minimum support values between 2 and 20. As can be seen in Figure 5, among all five, $D_{\text{MSDB-2}}$ performs the best and $D_{\text{MSDB-4}}$ the worst. Three of the five datasets achieve their highest accuracy when the number of bins is set to 13; however, $D_{\text{MSDB-4}}$ performs better at 19 bins, and $D_{\text{MSDB-5}}$ at 11 bins. Thus, we include the corresponding results in Figure 5.

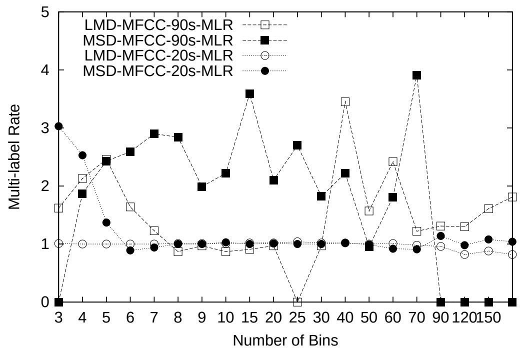


**Figure 9**. 5 genres of $D_{\text{LMD}}$ and $D_{\text{MSDB-5}}$ at $FRQ = 95$.

We observe that all three evaluation measures, $recall$, $precision$, and $accuracy$, obtain very similar values to each other in our experiments, as can be seen in Figures 6 and 7. It can also be seen in Figures 2 through 7, that AMGC does not behave arbitrarily, when given different datasets or different parameter settings. This confirms that our approach is stable and concludes our work towards $G_2$.

During our experiments, we notice that for some values of minimum support and for some numbers of bins, AMGC performs much better than choosing genre assignment at random. However, with other values of these parameters, AMGC predicts majority of music to be of one genre. Moreover, sometimes it votes for all genres equally, where MLR becomes equal to the number of genres. Furthermore, we encountered certain parameter settings, when some or all genres were not represented by any classification rules. We investigate the behaviour of MLR and the number of genres present in both $D_{\text{LMD}}$ and $D_{\text{MSDB}}$ through further experiments and report our findings in Figures 8 and 9. Here, we set the minimum support to 20 and then to 90 for both datasets. As can be seen in Figure 8, at the higher minimum support, some genres are discarded, due to removal of intersections during Stage 2 of our approach. Meanwhile, Figure 9 illustrates that AMGC behaves as a single label classifier, because we remove rules that are found among any genre-pair, thus, the remaining rules are representative of a single genre.

When experimenting with our approach on music genre classification using different features in $D_{\text{MSDB}}$, we use the same genre assignment and alternate the features. This helps us confirm that difference in content-based features result in different classification performance. Hence, different features are more or less useful for the genre classification task, which is reflected by the *feature selection* task in MIR.

In our experiments, we notice that it may take a long time to pre-process the data and train the classifier. However, the resulting classification model is very fast, where its speed can be expressed as the number of classification rules multiplied by the number of music pieces to be classified.

## 5. CONCLUSION

In this paper, we introduce a novel approach to MIR, namely, using association analysis to help music genre classification. Association analysis looks for frequent patterns in music data, which represent the similarity of all music pieces in a given genre.

Through experiments, we demonstrate the effectiveness of our approach and confirm that association analysis can be applied to music data. However, there is still room for improvement, which includes feature extraction, feature selection and discretization. We believe that as they improve, our method will also improve. We can also take some immediate steps to improve our classifier by tuning the two parameters, minimum support for mining frequent items and the number of discretization bins. Our experiments demonstrate that these two parameters are directly related to the performance of our classifier, and they vary depending on the data. Hence, tuning these parameters to each specific dataset will improve the classification accuracy. We leave these to our future work.

## 6. REFERENCES


[1] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, volume 22, pages 207–216. ACM, 1993.

[2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, volume 1215, pages 487–499. Morgan Kaufmann Publishers Inc., 1994.

[3] A. Anglade, R. Ramirez, and S. Dixon. Genre classification using harmony rules induced from automatic chord transcriptions. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 669–674. ISMIR, 2009.

[4] T. Arjannikov, C. Sanden, and J. Z. Zhang. Verifying tag annotations through association analysis. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 195–200. ISMIR, 2013.

[5] C. DeCoro, Z. Barutcuoglu, and R. Fiebrink. Bayesian aggregation for hierarchical genre classification. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 77–80. ISMIR, 2007.

[6] Z. Fu, G. Lu, K. M. Ting, and D. Zhang. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, 13(2):303–319, 2011.

[7] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., the second edition, 2006.

[8] F.-F. Kuo, M.-F. Chiang, M.-K. Shan, and S.-Y. Lee. Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, pages 507–510. ACM, 2005.

[9] M. Li and R. Sleep. Genre classification via an lz78-based string kernel. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 252–259. ISMIR, 2005.

[10] T. Li, O. Mitsunori, and G. Tzanetakis, editors. *Music Data Mining*. CRC Press, 2012.

[11] T. Li, M. Ogihara, and Q. Li. A comparative study on content-based music genre classification. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 282–289. ACM, 2003.

[12] C. Liao, P. Wang, and Y. Zhang. Mining association patterns between music and video clips in professional mtv. In B. Huet, A. Smeaton, K. Mayer-Patel, and Y. Avrithis, editors, *Advances in Multimedia Modelling*, volume 5371 of *Lecture Notes in Computer Science*, pages 401–412. Springer Berlin Heidelberg, 2009.

[13] A. Meng and J. Shawe-Taylor. An investigation of feature models for music genre classification using the support vector classifier. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 604–609. ISMIR, 2005.

[14] K. Neubarth, I. Goienetxea, C. Johnson, and D. Conklin. Association mining of folk music genres and toponyms. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 7–12. ISMIR, 2012.

[15] S. Russel and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson education Inc., the third edition, 2010.

[16] A. Schindler, R. Mayer, and A. Rauber. Facilitating comprehensive benchmarking experiments on the million song dataset. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 469–474. ISMIR, 2012.

[17] C. N. J. Silla, C. A. A. Kaestner, and A. L. Koerich. The latin music database. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 451–456. ISMIR, 2008.

[18] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, July 2002.

[19] L. Xiao, A. Tian, W. Li, and J. Zhou. Using a statistic model to capture the association between timber and perceived tempo. In *Proceedings of the International Society for Music Information Retrieval*, pages 659–662. ISMIR, 2008.