MODIFIED AIS-BASED CLASSIFIER FOR MUSIC GENRE CLASSIFICATION

Noor Azilah Draman, Campbell Wilson and Sea Ling
Caulfield School of IT,
Faculty of Information Technology, Monash University,
Caulfield East, Melbourne, Victoria, 3145
nadra1@student.monash.edu.au
campbell.wilson@infotech.monash.edu.au
chris.ling@infotech.monash.edu.au

ABSTRACT

Automating human capabilities for classifying different genre of songs is a difficult task. This has led to various studies that focused on finding solutions to solve this problem. Analyzing music contents (often referred as content-based analysis) is one of many ways to identify and group similar songs together. Various music contents, for example beat, pitch, timbral and many others were used and analyzed to represent the music. To be able to manipulate these content representations for recognition: feature extraction and classification are two major focuses of investigation in this area. Though various classification techniques proposed so far, we are introducing yet another one. The objective of this paper is to introduce a possible new technique in the Artificial Immune System (AIS) domain called a modified immune classifier (MIC) for music genre classification. MIC is the newest version of Negative Selection Algorithm (NSA) where it stresses the self and non-self cells recognition and a complementary process for generating detectors. The discussion will detail out the MIC procedures applied and the modified part in solving the classification problem. At the end, the results of proposed framework will be presented, discussed and directions for future work are given.

1. INTRODUCTION

Music genre is defined as classes or groups of songs that categorizes a collection of songs that have similar characteristics. It is a label created by music experts so that these songs are easily described and recognized [1]. There have been various studies on music genre classification over the years where generally the focuses would be on the type of features extracted, feature extraction techniques, feature selection mechanisms, and feature classification algorithms. This is because music genre classification is a unique topic, and an investigation that tries to imitate human capability to identify music. It is a process to automate the human skills in recognizing and grouping different type of music into categories by using their hearing senses and logical judgment.

Our research is also about automating the human identification process where we are investigating an algorithm from Artificial Immune System (AIS), called the modified immune classifier (MIC). MIC is a modification of negative selection algorithm, introduced in writer identification study [2]. Negative selection algorithm is one of a few algorithms developed in AIS domain where it stresses the antigen recognition process. Two processes involved: monitoring, a process of recognizing self/non-self cells by performing the affinity binding, and censoring, the process where antibodies (also known as detectors) are randomly generated to match with the antigens. The recognized antigens are called self cells whereas the non-recognized antigens are known as non-self cells. In the human immune system, recognized antigen is referring to cells that prevent human body from disease and non-recognized antigens are referring to cells that bring diseases to human body. MIC eliminates the process to generate detectors randomly, which is the main aspect of the NSA, by introducing a complementary process. This complementary process will define self cells based on how many classes of data they need to identify and then generate the detectors accordingly.

However, to be able to apply the modified immune classifier in this research, which is to be able to identify and recognize different groups of music genre, we need to change some part of the classifier in order to achieve high accuracy of results. We will discuss the changes that we have made later.

We present this paper with the intention of discussing music genre classification that applies modified immune classifier in the classification process. We are discussing in detail the feature extraction and feature selection processes except to explain the features used in the experimental work and the techniques used to select relevant and significant features. We elaborate the AIS approach in the context of music genre classification, their consequences in music recognition performances whether the approach will have a major impact to the classification performances.

We organize the remainder of this paper as follows: Section 2 discusses previous research in music genre recognition. Section 3 discusses the MIC and the changes part, the censoring and monitoring stages, and how these stages relate to the feature extraction, selection, and classification. Section 4 then will be discussing the experimental setup and the classification results. We outline some concluding remarks in the last section.

2. BACKGROUND OF STUDY

In the music genre identification and classification studies, initiated research was to solve problems that occur during recognition such as, deciding which song belongs to which genre. [3], for example, did an early work of classifying songs into different categories of genre using human auditory skills. Since then, many studies to find solutions to increase the automation performances occurred. Various recorded attempts to solve this problem are in [2] – [9]. Not only the problem of automating the process of classification but the question of how to fill the gap of accuracy behind human skilled classification [3] also need to be answered and solved.

[1] contributed by introducing new music features from pitch, timbre and rhythm contents. Their experiments on genre classification have shown that their attempts can be investigated further as the classification accuracy results were around 61 percent only. The focus of their research was to introduce a new range of music features for music genre classification. As the extracted features are too numerous, many irrelevant and insignificant features were used in their experiments that contributed to the low level of performances.

[10] introduced a new technique to extract music features called Daubechies Wavelet Coefficient Histograms (DWCHs) with a purpose to overcome the classification accuracy problems in the previous study. The authors used the Daubechies wavelet filter, *Db8*, to decompose music signals into layers where at the end of each layer they constructed histograms of coefficient wavelet. During experiments they combined their new feature with [1] features and improved the results but not by much.

There is also another attempt that used pitch, rhythm and timbre contents to classify music into different genres [11]. In this study, the author used the neural network based classifier which was not tested in the previous two studies. Again similar problem that related to the classification performance occurred. The experiments have shown that the accuracy was quite high when the classification processes were to recognize one or two genres only. But, as the classes of genres increased, the performances began to decrease.

[12] proposed a solution to the problem mentioned above. The authors proposed a new feature extraction method called *InMAF.* This new method was quite different from previous approaches where previously, they relied mostly on the spectrum characteristics of music content. *InMAF* on the other hand integrated the acoustic features and the human musical perception into music feature vectors to increase the classification performances. The classification results were so impressive that the achieved accuracies were as high as ninety percent. However, these outcomes were the results of a combination of this new method with pitch, rhythm and pitch contents. There is no classification result from any individual features recorded in the study.

[8] attempted to classify the music genre using MIDI (Musical Instrument Digital Interface) and audio features, such as pitch, rhythm and timbre features by using the data from [13], which contained two different sets of features, the first was MIDI features and the other group was the audio features. However the attempt was not that successful as the result did not show any major improvement in the classification performances.

A new recent study proposed a new approach to classify music genre by emphasizing the features from cepstral contents, such as MFCCs, OSC and MPEG 7 representations [14]. They introduced a novel set of features that were derived from modulation spectral analysis of the spectral representations, and these features were the *Mel-Frequency Cepstral Coefficients* (MFCC), *Octave-based Spectral Contrast* (OSC), *Normalized Audio Spectral Envelope* (NASE) and *Modulation Spectral Analysis* of MFCC, OSC and NASE. Their experiments were conducted on individual features and combinations of features.

The results were very good, where the combination of features tested were able to achieve the accuracy around twenty percent higher than any studies that we have discussed so far. That was an impressive achievement since low classification accuracy is the major problem faced by the domain.

3. AIS-BASED CLASSIFIER

In this part, we discuss Artificial Immune System (AIS) approach specifically on the modified negative selection algorithm (MIC) to classify the music genre. According to [15], the human immunology system inspired this domain to observe the immune functions, models, and principles of immunology. Some references on AIS-based classification task can be found in [16 -17].

AIS are adaptive systems, emulating human body immunology system to solve problems. It is concerned with abstracting the whole concept of immune system to computational systems in solving problems from mathematics, engineering, and information technology point of view. AIS is developed based upon a set of general purposes algorithms that are modelled to generate artificial components of the human immune system. [15] defined AIS as an adaptive system which is enthused by biological immunology and observed functions, principles and models to problem solving.

[18] introduced negative selection algorithm as inspired by negative selection of T-cells in thymus. The algorithm focused on recognizing self or non-self cells where it eliminated the T-cells which thymus does not recognized. Detail explanations of how negative selection algorithm works is in [19]. As has been investigated before, it would be impossible to apply NSA without modification as each problem and solutions are different. However, we will not discuss the NSA further as it is not in the research scope.

In the next section, we will discuss the MIC, the censoring and monitoring stages including features conversion, complementary and identification processes that we have applied to suit with the problem in hand. Then we continue the discussion with detailed explanation of the changes that we have made in the identification accuracy calculation.

3.1 Modified Immune Classifier (MIC)

The inspiration to investigate MIC in this research comes from a writer identification study [2] where the proposed classifier to identify different writers has provided excellent results as the identification test achieved the accuracy as high as 99 percent. The recognition is evaluated by emphasizing the affinity binding or similarities between those cells.

In this new version of NSA, the author introduced a complementary process, which is a process of generating detectors to detect antigens. Originally, in NSA, the detectors are randomly generated and cost some time. They also do not contain enough information to recognize the whole range of antigens.

This would become a problem because, in order to recognize the antigens, the generated detectors shall not be created randomly as the process will not guarantee there will be enough detectors. By having the complementary process, the detectors will be generated accordingly to compensate the antigens as has been done in the writer identification research where the complementary process generated detectors according to the number of writers that should be recognized. Imitating the immune system's function, MIC works:

(1) self-cells (feature vectors) are transformed into antibodies (detectors) to detect similar cells (antigen),
(2) during detection (identification) antibodies will do the affinity binding with the antigens (finding similarities),
(3) both cells will bind (matched) if there are similarities occurred – antibodies detected antigens as similar to it cells – a pattern is recognized

As has been mentioned earlier, censoring and monitoring modules are two important process of MIC. We will discuss them next.

3.2 Censoring and monitoring modules

Censoring module is responsible to produce detectors, which is the key aspect of identification. This module normally starts after feature extraction and feature selection processes. It involves data feature conversion where the features will be represented by binary bit strings (for example, a feature vector, -3.4523123 is converted into a binary string, 101011001 using –XOR operation). After the conversion, the binary bit strings then will go through the complementary process and become the detectors.

We applied the supervised learning experiments in this research and we used training data to generate the detectors. Once generated, we used them in the classification process by comparing the detectors and generated antigens (we converted testing data into antigens). The comparison occurred in the monitoring module (the training model/detectors created earlier to predict the testing data/antigens) and it was to find matched data between detectors and antigens. If matched, we then calculate the affinity binding.

The comparison produced binary bit '1' or '0' where bit '1' means the data is bind. However, in this scenario, we will use the word 'match' instead of 'bind' to define the similarities. Figure 3.1 illustrates both modules where two important things occurred in censoring module, which are the conversion data from feature vectors into binary bit strings using –XOR and detectors generated processes. In monitoring module, two important things also occurred, which are antigens generated from testing data and identification processes. During binary matching process, we used Hamming distance technique to calculate matched binary bits.

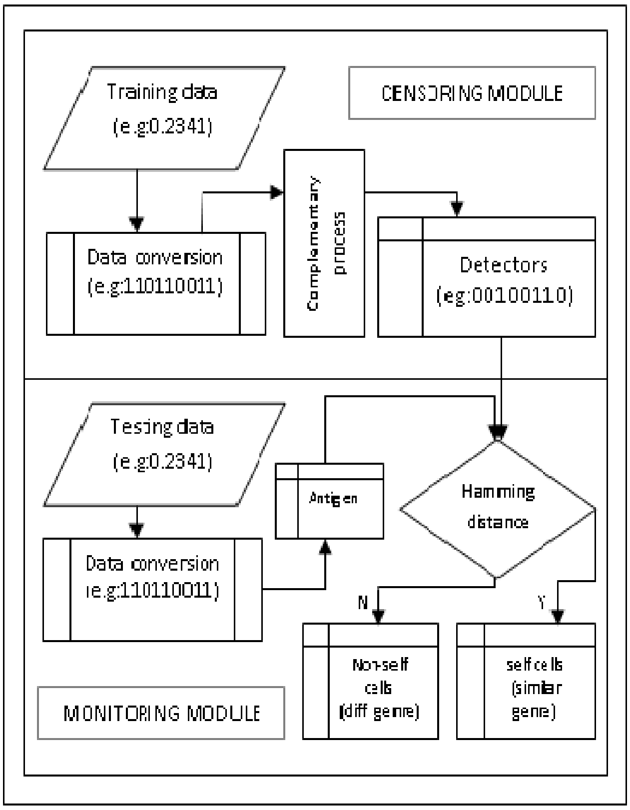

Figure 1. Censoring and monitoring modules

3.3 Accuracy calculation

In the writer identification problem, the calculation emphasized the recognition of each feature where these features will be calculated individually based on a threshold value. The accuracy would be based on how many features were correctly classified. To apply MIC to our problem, we concentrated on the threshold value in the accuracy calculation where the value will be our benchmark to decide whether the songs are classified accurately or not.

During the process, we calculated the result first by combining all features and produced the data accuracy percentage. Then we compared the accumulated value with the threshold value percentage. If the percentage of the combined features is higher than the threshold value, the data then is labeled as accurately classified. The following Table 3.1 and 3.2 will show the difference between the writer identification calculation and ours.

The difference between the original MIC proposed in [2] with ours is that we combined all the feature vectors as one whole data and calculates the matched bits before we compare them with the threshold value, whereas in the author identification study, the matched bit is calculated separately for each feature and the accuracy is computed based on the total amount of features that exceeded the threshold value.

Category	Calculation formulas
Feature matching stage	Num_of_bit_match ≥ threshold
Image accuracy stage	(Num_of_feature_match / num_of_feature) × 100
Data accuracy stage	(Num_of_genre_match / num_of_testing_data) × 100

Table 1. The writer identification calculation

Category	Calculation formulas
Data genre accuracy %	∑ bits_matched / ∑ features_bits × 100
Threshold (***r***) %	(∑ ***r*** * num_of_features / ∑ bits_per_feature * num_of_features) × 100
Dataset accuracy	(Num_of_genre_match / num_of_testing_data) × 100

Table 2. The music genre accuracy calculation

4. EXPERIMENTS

In this section, we explain our conducted experiments to evaluate the proposed algorithm**.**

4.1 Datasets

We used Latin music datasets which contains 3160 music pieces in MP3 format classified in 10 different musical genres [20][21]. The songs were grouped into 10 genres: Tango, Bolero, Batchata, Salsa, Merengue, Axé, Forró, Sertaneja, Gaúcha and Pagode. The extracted music features were from timbre contents (containing MFCC, spectral centroid, roll-off, flux, time domain zero crossings), pitch-histograms related features and beat calculated features. The features were extracted using MARSYAS [22] where the combined total of the features produced 30 vectors for each song.

We have prepared training and testing datasets where similar data is used in the experiments except that the data for WEKA experiments was in the attribute related file format (ARFF) and in the data file (DAT) format for MIC demonstrations.

4.2 Feature selection technique

We have used WEKA tool to select relevant and significant features. We used filter approach in this study because it is more practical and suitable for our problem as the approach is independent and work separately from the classifier. The filter approach also works faster than wrapper and embedded approaches.

We have selected significant features using two search approaches, which are the best first search algorithm and the greedy hill search algorithm. The techniques that we used to do the best first search selection and the greedy hill selection are the FilterSubsetEval, the CFSSubsetEval and the ConsistencySubsetEval. The produced selected features from these techniques contained 13, 17, and 18 feature vectors.

We tested the MIC algorithm in the classification processes by defining the threshold value as 12. The reason is that we want to compare the proposed MIC with other classifiers without evaluating various threshold values to select the best one. The chosen threshold value is considered practical and enough to determine the reliability of the proposed technique.

Table 3 describes the feature vectors in detail where they have been numbered (1 to 30) for easy identification.

Features	Description
1 - 6	Beat-related features (peak histograms, amplitude and period)
7 - 25	Timbral features (mean and standard deviation of spectral centroid, rolloff, flux, zero crossings, MFCC, low energy)
26 - 30	Pitch related features (folded and unfolded histograms, period, amplitude pitch interval of unfolded histograms)

Table 3. Features description

4.2 Classification

For comparison purposes, we used classifiers from Waikato Environment for Knowledge Analysis (WEKA) [23] and the MIC algorithm that we have built using C++ language. We have chosen few classifiers from different category in WEKA.

We have setup the experiment cases according to the selected features from selection process. We also have setup experiments to test individual group of features and combinations between the groups. The reason is that we want to test the robustness of our program and the reliability of AIS-based classifier performance in our classification problems. Table 4, 5, and 6 will explain these cases in detail.

Cases	Description
C1	Features 1, 2, 6, 9,10, 13, 17, 18, 22, 25, 26, 28
C2	Features 1, 4, 6, 7, 9, 10, 12, 13, 14, 15, 16, 17, 18, 21, 23, 26, 28
C3	Features 1, 4, 6, 9, 10,11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 25, 26
C4	Contains all 30features

Table 4. List of selected features

Cases	Description
F1	Features 1 – 6 (beat related features only)
F2	Features 7- 25 (timbral related features only)
F3	Features 26 – 30 (pitch related features only)

Table 5. Individual group of features

Cases	Description
FBP	Combination of beat and pitch related features
FBT	Combination of beat and timbral related features
FTP	Combination of timbral and pitch related features

Table 6. Combination of group features

4.3 Results

Table 7, Table 8, and Table 9 list all classification results that we have obtained from the prepared cases classification experiments.

Technique Cases	Case 1	Case 2	Case 3	Case 4
BayesNet	50.00%	53.00%	53.00%	58.33%
SMO	46.00%	48.67%	57.00%	56.33%
IB1	49.00%	51.67%	56.00%	57.00%
Bagging	42.33%	44.00%	47.67%	48.33%
J48	38.00%	38.33%	42.00%	42.00%
MIC	**99.33%**	**95.00%**	**92.67%**	**73.00%**

Table 7. Selected features cases

Technique Cases	F1	F2	F3
BayesNet	29.67%	29.67%	49.33%
SMO	27.33%	31.33%	50.00%
IB1	27.33%	**100 %**	50.33%
Bagging	30.33%	66.33%	53.00%
J48	28.67%	72.00%	45.00%
MIC	**100 %**	**100 %**	**93.33%**

Table 8. Individual group of features cases

Technique Cases	FBP	FBT	FTP
BayesNet	39.3333%	53.0000%	54.3333%
SMO	33.3333%	57.6667%	56.3333%
IB1	38.3333%	55.3333%	56.0000%
Bagging	35.0000%	49.3333%	52.6667%
J48	37.3333%	40.3333%	48.3333%
MIC	**99.00%**	**79.33%**	**91.33%**

Table 9. Combination of group features cases

In Table 7, for feature selection cases, all cases except for the data without feature selection, MIC has obtained the accuracies over 90% compared to other classifiers. The performances of other classifiers did not show any significant improvement compared to MIC.

Table 8, which is referring to the individual group of features experiments. Overall performances for each feature when tested with various classifiers have shown that beat related features produced the lowest accuracy results. Timbral related features came in second however, when tested with MIC classifier pitch and timbral features produced similar percentages. Bagging classifier also produced similar result when tested the timbral related features to classify the songs.

In Table 9, WEKA classifiers produced almost similar results when we experimented with both beat+timbral related and timbral+pitch related features. The lowest accuracy recorded with beat+pitch related features when these features were used for classification. However, the opposite case occurred when the data were classified using MIC classifier because the lowest accuracy recorded when beat+timbral related features were tested.

5. CONCLUSION

The availability of techniques and methods for classification in music analysis field proved that researchers in this area are very concerned with the performance. As the collections of digital songs keep increasing online, their studies have contributed a major breakthrough to the internet users and others.

In this paper, we have experimented and explained the proposed MIC in different category of cases. In each experiment, MIC has outperformed almost every classifier except for Bagging technique where in one of the cases, the result is exactly similar to what MIC has produced. The obtained results have clearly shown that MIC is a new prospective approach for music genre classification. It has been proven the proposed classifier in music recognition research has surpassed other classifiers and the improvement of classification accuracy is phenomenal. The results also showed that among the features, timbral has provided us good classification result in the most cases except for the combined features cases.

We strongly believe that our discussion throughout this paper has given opportunities to other researchers in this area of studies to fill the gaps, to explore further and to provide solutions to the known and un-known problem that has yet to be discovered. Future work will include an investigation on how to manage efficiently the threshold value and probably later on, exhaustive search approach should be applied to evaluate the highest threshold value that can provide high classification accuracies.

6. REFERENCES

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals", IEEE Transactions on Speech and Audio Processing, vol.10, no.5, pp. 293-302, 2002.

[2] Draman, A. K. "Authorship invarianceness for writer identification using invariant discretiation and modified immune classifier", PhD thesis. University of Technology Malaysia, 2009.

[3] Lippens, S., Martens, J. P. & Mulder, T. D. "A comparison of human and automatic musical genre classification" *Acoustics, Speech, and Signal Processing,* 2004.

[4] Brecheisen, S., Kriegel, H. P., Kunath, P. & Pryakhin, A. "Hierarchical genre classification for large music collections". *2006 IEEE International*

Conference on Multimedia and Expo - ICME 2006, Vols 1-5, Proceedings, 1385-1388, 2006.

[5] Ahrendt, P., Larsen, J. & Goutte, C. "Co-occurrence models in music genre classification". *2005 IEEE Workshop on Machine Learning for Signal Processing (MLSP)***,** 247-252, 2005.

[6] Bagci, U. & Erzin, E. "Boosting classifiers for music genre classification". *Computer and Information Sciences - Iscis 2005*, Proceedings, 3733, 575-584, 2005.

[7] Bagci, U. & Erzin, E. "Inter genre similarity modeling for automatic music genre classification". *2006 IEEE 14th Signal Processing and Communications Applications*, Vols 1 and 2**,** 639-642, 2006

[8] Cataltepe, Z., Yaslan, Y. & Sonmez, A. "Music genre classification using MIDI and audio features", *Eurasip Journal on Advances in Signal Processing*, 2007.

[9] Cheng, H. T., Yang, Y. H., Lin, Y. C., Liao, I. B. & Chen, H. H. "Automatic Chord Recognition for Music Classification and Retrieval". *Ieee International Conference on Multimedia and Expo,* Vols 1-4**,** 1505-1508, 2008.

[10] T. Li, M. Ogihara and Q. Li, "A comparative study on content-based music genre classification", *Proceedings of the 26th annual international ACM SIGIR*, Toronto, Canada, pp. 282-289, 2003.

[11] R. Neumayer and A Rauber, "Integration of text and audio features for genre classification in music information retrieval", In *Proceeding of 29th European Conference on Information Retrieval, Rome, Italy,* pp. 724-727, 2007.

[12] Shen, J. L., Shepherd, J. & Ngu, A. H. H. "On efficient music genre classification". Database Systems for Advanced Applications, Proceedings, *3453, 253-264.*

[13] Mckay, C. & Fujinaga, I. "Musical genre classification: Is it worth pursuing and how can it be improved*". ISMIR 2006*. Victoria Canada.

[14] Lee, C. H., Shih, J. L., Yu, K. M. & Lin, H. S. "Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features". *Ieee Transactions on Multimedia,* 11, 670-682, 2009.

[15] L.N. de Casto and J. Timmis, "*Artificial immune system: A new computational intelligence approach",* Great Britain, Springer, pp. 76-79, 2002.

[16] S. Doraisamy, S. Golzari, N. M. Norowi, M. N. Sulaiman and N. I. Udzir, "A study on feature selection and classification techniques for automatic genre classification of traditional Malay music", in *Proceedings of Ninth International Conference on Music Information Retrieval (ISMIR'08)*, Philadelphia, Pennsylvania USA, pp. 331-336, 2008.

[17] D.N. Sotiropoulus, A.S. Lampropoulus and G.A. Tsihrintzis, "Artificial immune system-based music genre classification", in *New Directions in Intelligent Interactive Multimedia*, pp. 191-200, 2008.

[18] R.-B. Xiao, L. Wang and Y. Liu, "A framework of AIS based pattern classification and matching for engineering creative design", in *Proceedings of the First International Conference on Machine Learning and Cybernetics*, Beijing, China, pp. 1554-1558, 2002.

[19] S. Forrest, A.S. Perelson, L. Allen and R. Cherukuri, "Self-nonself discrimination in a computer", in *Proceedings of 1994 IEEE Computer Society Symposium on Research in Security and Privacy*, Oakland, CA, USA, pp. 202-212, 1994.

[20] C.N. Silla, A.L. Koerich and C.A.A. Kaestner," The Latin Music Database"*Proceedings of Ninth International Conference on Music Information Retrieval (ISMIR'08)*, Philadelphia, Pennsylvania USA, 2008, pp. 331-336.

[21] C.N Silla, A.L Koerich ans C.A.A Kaesner. "A feature selection approach for music genre classification". *International Journal of Semantic Computing,* Vol 3, No 2, 183 -208, 2009.

[22] G. Tzanetakis, P.Cook, "MARSYAS: A Framework for Audio Analysis". *Organized Sound Journal.* Vol 4 (3), 2000.

[23] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, "The WEKA Data Mining Software: An Update"; *SIGKDD Explorations*, Volume 11, Issue 1, 2009.