World Academy of Science, Engineering and Technology
International Journal of Electrical and Information Engineering
Vol:10, No:2, 2016

Open Science Index, Electrical and Information Engineering Vol:10, No:2, 2016 publications.waset.org/10006934/pdf

Genetic Algorithms for Feature Generation in the Context of Audio Classification

José A. Menezes, Giordano Cabral, Bruno T. Gomes

***Abstract*—**Choosing good features is an essential part of machine learning. Recent techniques aim to automate this process. For instance, feature learning intends to learn the transformation of raw data into a useful representation to machine learning tasks. In automatic audio classification tasks, this is interesting since the audio, usually complex information, needs to be transformed into a computationally convenient input to process. Another technique tries to generate features by searching a feature space. Genetic algorithms, for instance, have being used to generate audio features by combining or modifying them. We find this approach particularly interesting and, despite the undeniable advances of feature learning approaches, we wanted to take a step forward in the use of genetic algorithms to find audio features, combining them with more conventional methods, like PCA, and inserting search control mechanisms, such as constraints over a confusion matrix. This work presents the results obtained on particular audio classification problems.

***Keywords*—**Feature generation, feature learning, genetic algorithm, music information retrieval.

I. INTRODUCTION

A good choice of features is an essential part of most machine learning algorithms. This is particularly true for automatic audio classification, which is our area of interest, both for musical and non-musical applications. Traditionally, this is a hand-crafted activity, requiring human labor and relying on expert knowledge. Recent works, however, try to make the discovery of pertinent features an automatic and expert-independent task.

Feature learning techniques [8], for instance, intends to learn a transformation of raw data to a representation which could be used in machine learning tasks. For example, which are bringing major advances to the music information retrieval community, and has been gaining momentum in last years.

Some solutions, however, rely on an explicit search over the audio feature space as their meta-learning approach. These solutions have the advantage of using, as a starting point or as vocabulary, a list of the most frequent features in the domain (for example, acoustic features). Thus, these solutions can work in a less agnostic way.

For audio, there is the work from Zils and Pachet with the Extractor Discovery System (EDS) [4] at SONY CSL Paris, and McKay's jMIR suite [2], which includes the automatic classification engine (ACE) [3].

jMIR's ACE uses meta-learning for selecting, optimizing and applying machine learning algorithms to music research. The user can compute and select audio features, as well as finding new ones via feature learning.

SONY's EDS, on the other hand, along with its many useful functionalities, such as feature selection, optimization of classifiers, and visualization of information, relied on a crucial step of searching the space of analytical features to find the most relevant ones to a particular problem. This search used a genetic approach, where each feature was considered an individual, and each mathematical operation was considered a gene.

We found this approach particularly interesting and, despite the undeniable advances of feature learning approaches, we wanted to go a step further in the use of genetic algorithms to find audio features. That is the subject of this work.

Section II presents state of the art. Section III present the proposed solution, in Section IV is presented the experiments and in Section V the conclusions and future works. Lastly, the acknowledgment and references are presented.

II. STATE OF THE ART

Audio classification normally has two very important steps: a) finding, processing and selecting features; these features will transform raw data into more meaningful information and b) running classification algorithms (such as k-nn, neural networks, support vector machines, etc.) which will take this processed information as input.

Finding good features, thus, have a huge impact on the overall quality of the final classifier. Traditionally, this was a responsibility of experts, which analyzed the particular audio classification task, and explored the use of different features that might seem pertinent to the problem in hands. But many techniques have appeared in the attempt of making the whole process automatic.

A. Feature Learning and Feature Generation

Although feature learning may be seen as a general term for any process of learning features to make it more mathematically or computationally compatible to a particular problem, it normally means particular transformations of the input data representation. Feature learning techniques are generally divided into two groups: supervised and unsupervised, given the existence or not of a ground truth. In the first group, we can find classical machine learning algorithms, such as neural networks, and in the second group we can find methods such as PCA [6] and matrix factorization [7].

J. A. Menezes is with the Statistic and Informatics Department, Rural Federal University of Pernambuco, Brazil (e-mail: menezes.jaam@gmail.com).

G. Cabral and B. T. Gomes are with the Informatics Center, Federal University of Pernambuco, Brazil (e-mail: grec@cin.ufpe.br, btmg@cin.ufpe.br).

Many categorizations, however, may arise. For example, the ones based on an agnostic approach, such as PCA, where the whole process is automatic and data were driven, and those allowing more control from the user, for example, based on explicit searches over the audio feature space. In this work, we are particularly interested in techniques capable of finding new features, not only selecting or adjusting parameters. And we want to have some level of control to induce a search on particular subspaces. In the lack of an established term, we call it here "feature generation".

B. Tools

During last years, some audio feature generation systems appeared. The Autonomous Classification Engine (ACE), from the McKay's jMIR suite, is an example. It allows the user to calculate and select features from a given list, to apply them dimensionality reduction techniques, which possibly results in new features. The jMIR is an open source and full-fledged suite. However, the generation of new features, only based on methods such as PCA, is quite limited.

The Extractor Discovery System (EDS), developed by Pachet and Zils at SONY CSL Paris uses a different approach. It is heavily based on genetic algorithms [1], where each feature is considered an individual and each mathematical operation are considered a gene. The system, thus, modifies or recombines features, generating completely new ones.

Indeed, as feature generation may be seen as a search problem, the use of evolutionary computation seems quite appropriate. A priori, the search space derived from all possible combination of operators is infinite, leading to an unlimited number of optimal solutions. However, it can be reduced, in practice, if the maximum number of operators is limited. Given a reasonable maximum number of operators, such as ten, the search space is still unviable to be explored by systematic searches, favoring the use of parallel and heuristic searches, such as the genetic ones.

However, EDS feature generation algorithm still uses a shallow and mono-objective approach, finding features solely by the individual's aptitude. The evolution of audio features follows the direction of those with higher fitness. In some cases, however, this may not be sufficient. For example, there may be other objectives beyond the number of classified instances. On the other hand, particular situations (such as a high number of false negatives) must be avoided, as it will be illustrated in next section.

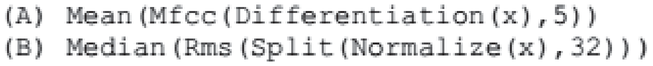
C. Constraints

These "situations" may be modeled as constraints. A constraint interferes in the fitness of an individual based on a confusion matrix. This matrix offers more information to indicate the performance of the classifier according to the nature of the problem. In most cases, the goal is to maximize the number of classifications in the blue area. However, some exceptions may appear. For example, in an audio-based security system, it is critical that the false negatives be completely avoided. Although we acknowledge that more sophisticated techniques, such as ROC curves, are useful, we care that the control of the evolution could be left to the user. This user can, thus, specify particular rules over the confusion matrix, to serve to restrict the search space.

III. PROPOSED SOLUTION

Given the approaches of jMIR/ACE and EDS, we wanted to go a step further, proposing a genetic algorithm to find audio features whose final classification could satisfy constraints over a confusion matrix.

We start by un finite set of elementary operators. For example, mathematical operations (addition, multiplication, mean, etc.), signal processing operations (Fourier transform, filters, spectral properties, etc.), music specific (pitch, chroma, etc.). These operators are written in a way that we can verify if recombinations or concatenations result in valid expressions or not (the output of an inner operation matches the input of an outer operation). Fig. 1 shows two audio features (representing two individuals in the genetic search).



```
(A) Mean(Mfcc(Differentiation(x),5))
(B) Median(Rms(Split(Normalize(x),32)))
```

Fig. 1 Example of individuals (A) Average of the 5 first cepstral coefficients of the derivative of the signal x. (B) Mean value (Median) of the energy (Rms) of successive frames (split) of 32 samples long in the normalized x [5]

This is exactly the same as EDS approach. The genetic mechanisms are also the same as in EDS, as described in [5]: cloning, mutation, deletion, addition, and crossover. We wanted to do so, to assure comparability with EDS, since it is a closed solution.

We had to define some other aspects as well: meta-parameters (number of generations, population size, etc.), the fitness formula (responsible for indicating whether a feature is better or worse than another is). For the last one, we ended by using a generic and fast classification algorithm, so the success rate obtained was used as the fitness of the feature.

As said before, despite the advantages of the EDS system, it lacks the specification of flexible objectives (for example, which part of the confusion matrix is more important for each problem), what we defend can be done with constraints.

Section IV.A analyses one of these situations: a security alarm which has to be triggered in case of violent events, such as gunshots. In this case, besides of enhancing the performance, algorithms should restrict the number of false negatives. In fact, not detecting an actual event (false negative) is a very important (almost unacceptable) error, while a false alarm (false positive) is widely acceptable.

We can, then, model our problem in the following way:

Objective function:

$$g(tp, tn, fp, fn) = max\,[(tp+tn)/(tp+fp+tn+fn)] \qquad (1)$$

Constraints:

$$fn = 0 \qquad (2)$$

or,

$$fn/total <= 0,05 \quad (3)$$

or, among other possibilities.

As can be seen in the example above a restriction operates on one of the confusion matrix variables (tp, tn, fp, and fn) and sets a percentage which the variable must correspond to a value lower, greater or equals to that threshold.

Thus, the fitness formula depends on the constraints of the problem in hands, and the individuals that do not satisfy these constraints are penalized during the genetic algorithm.

IV. EVALUATING

Our evaluations try to shed some light on two questions. First, that genetic algorithm is a more suitable approach to feature generation than dimensionality reduction techniques. Second, that the addition of control mechanisms in the genetic algorithm leads to more adapted solutions since they do not rely isolated on an error rate, precision, recall, accuracy, f1, or any other specific metric. With that, we want to show that there is still room for genetic algorithms to evolve, especially by adding multi-objective situations.

In next sections, we present the results obtained for three audio classification problems, with three techniques: meta-learning, using jMIR/ACE and PCA; a simple evolutionary algorithm, such as in EDS, and an enhanced evolutionary algorithm, including constraints.

A. Problem 1: Audio Alerta

The Audio Alerta is a system developed by Brazilian company D'Accord Music [8], [9] capable of recognizing dangerous situations like gunshots and car crashes via sound stream. The company agreed to us to use its database, consisting of almost 18,000 samples.

We performed an experiment where the classifier should be able to differentiate between gunshots and fireworks (very similar sounds). As explained before, the goal is not only to reduce the general error rate but to avoid at the maximum the appearance of false negatives. We performed the experiment on a subset of 62 samples.

Table I presents the final results. The genetic algorithm with constraints not only reached an overall maximum at 82.3% of success rate (against 79% of the best alternative), but it has only 11,3% of false negatives, the even of the alternatives, as it can be seen at the last line of the table. Even though 11,3% can still be considered a too high FN rate, the GA with the constraint is the one which reduces it the most.

TABLE I
PROBLEM 1: AUDIO ALERTA

Method	ACE + PCA	Simple GA	Ga with Constraint
Success Rate	79 %	82,3%	82,3%
FN Rate	11,3 %	11,3%	8,3%

Results for Audio Alerta base

B. Problem 2: Voice Recognition

This is a speaker identification problem, such as in access control of facilities. 80 samples have been used to identify whether the voice comes from an authorized person or not. In this case, we are not only interested in reducing the error rate, but the number of false positives (where access would be granted to an unauthorized person).

TABLE II
PROBLEM 2: VOICE RECOGNITION

Method	ACE + PCA	Simple GA	Ga with Constraint
Success Rate	93,7 %	76,25 %	76,25 %
FN Rate	2,5 %	17,5 %	17,5 %

Results for voice recognition base jMIR/ACE achieves best results both for overall performance and in constraint satisfaction

C. Problem 3: Nasal and Orals

This is a database with 60 examples of pronunciations of vowels with and without nasal phonemes. Although the evaluation of success, in this case, relies basically on the error/success rate, we decided to include constraints to prevent false positives and false negatives not to pass 25%.

TABLE III
PROBLEM 3: NASAL AND ORALS

Method	ACE + PCA	Simple GA	Ga with Constraint
Success Rate	68,3 %	75 %	68,3 %
FP Rate	18,3 %	8,3 %	15 %
FN Rate	13,4 %	16,7 %	16,7 %

Results for nasals and orals phonemes

Table III presents the results. As the sum of the false positives and false negatives represent the error rate, it is obvious its intrinsic relationship with the success rate. When the success rate is higher, the FP and FN are smaller, and vice-versa. Besides, FP is always greater than FN, disregarding the algorithm. When the success rate increases, FP and FN improves as well, so this kind of constraint (directly correlated to the success rate itself) has proven unusual.

V. CONCLUSIONS AND FUTURE WORKS

This work tried to analyze algorithms for feature generation, proposing enhancements and comparing them. Concretely, we wanted to take a step further in the genetic algorithm approach, as initiated by the EDS system, and compare it with a more conventional PCA-based approach. Results are inconclusive because:

1. Because the experiments only tested three audio classifications datasets.
2. because these datasets were small.
3. because each experiment had a different behavior.

Still, we can draw some conclusions. The first one is that there is room for improvements in genetic algorithms, especially regarding multi-objective problems. This is a very common situation, and this work presented a potential solution for it: the use of constraints. Despite the fact that we used it very lightly, it can be a major reason to prefer genetic algorithms in detriment of black box feature learning mechanisms, such as PCAs.

Results also show that the use of constraints improved the

Open Science Index, Electrical and Information Engineering Vol:10, No:2, 2016 publications.waset.org/10006934/pdf

overall success rates. This is probably due to the fact that the constraints specified were based on the error part pf the confusion matrix, ultimately reinforcing the importance of the success rate. A full multi-objective solution, with much more flexible constraints, could obviously lead to worse overall results, since constraints might contradict the main metric (in this case, the success rate).

Future works point to developing an actual multi-objective solution, probably inheriting the constraint mechanism. We also plan to repeat this kind of comparison with more datasets, each one comprising more samples, probably using some standard MIR datasets, such as those used in the MIREX competition. The use of more operators is also in the pipeline since the current solution only used a subset of operators. Adding new operators will enhance the expressivity of the algorithm, increasing the search space, and probably achieving better results. However, we forced the same limitation of operators to be used in every solution implemented; it may affect the comparison between PCA based and genetic algorithm based solutions.

ACKNOWLEDGMENT

We want to thank Daccord Music Software for providing us a dataset for this study. We also thank FACEPE and INES for supporting this research.

REFERENCES

[1] A. E. Eiben and J. E. Smith: *Introduction to Evolutionary Computing*, Springer, Amsterdam, 2003.
[2] C. McKay: “Automatic music classification with jMIR”. *Ph.D. Thesis*. McGill University, Canada, 2010.
[3] C. McKay, R. Fiebrink, D. McEnnis, B. Li, and I. Fujinaga: “ACE: A framework for optimizing music classification.” *Proc. of the International Conference on Music Information Retrieval*. 42–9, 2005.
[4] F. Pachet and A. Zils: “Evolving Automatically High-Level Music Descriptors from Acoustic Signals.” *Proc. of the International Symp. on Computer Music Modeling and Retrieval*. Springer Verlag LNCS, 2771, 2003.
[5] F. Pachet and P. Roy: “Exploring Billions of audio features.” *Proc. of the International Workshop on Content-Based Multimedia Indexing*, pp. 227 - 235, 2007.
[6] I.T. Jolliffe: *Principal Component Analysis*, Springer, Nova York, 2002.
[7] N. Srebro, J. Rennie; T. Jaakkola: “Maximum-Margin Matrix Factorization”, *Proc. of the Conference on Neural Information Processing System*, 2004.
[8] Y. Bengio, A. Courville, P. Vincent: “Representation Learning: A Review and New Perspectives”. IEEE Trans. PAMI, special issue Learning Deep Architectures, 2013.
[9] http://www.daccord.com.br/