2009 International Conference on Machine Learning and Computing
IPCSIT vol.3 (2011) © (2011) IACSIT Press, Singapore

Optimizing ANN's Architecture for Audio Music Genre Classification

Panagiotis S. Stergiopoulos and Odysseas B. Efremides [+]

University of Hertfordshire, Hatfield, UK,in collaboration with IST Studies
Dept. of Computer Science 72 Pireos Str., 183 46, Moschato, Athens, Greece

Abstract. The impact of the architecture of a neural network performing automatic audio music genre classification on accuracy it achieves is investigated herein. One evolutionary and three traditional optimization approaches proposing neural network architectures are compared based on the resulted classification accuracy. A value of 81% classification accuracy obtained by a neural network built as proposed by the genetic algorithm approach. This result is claimed to be satisfactory taking into account that no other optimizations, except for the architecture, have been applied to the neural network.

Keywords: music genre classification, neural network, genetic algorithm, optimization.

1. Introduction

During the last decade the amount of music stored in personal computer databases has been enormously grown. Thus, systems that deal with musical databases gained importance while the demand for Music Information Retrieval (MIR) applications keeps increasing [1]. Key issue of MIR applications is the automatic analysis of the contents stored in those musical databases. Unfortunately, most of the currently existed databases are indexed by text data, such as the song's or the artist's name [2]. This kind of indexing may be useful for classifying audio pieces according to the singer's name or the band, but is inefficient in classifying pieces into musical genres. The problem arises due to the fact that there are no strict boundaries between musical genres and there is no complete agreement in their definition. Members of any particular music genre share discrete audio characteristics (rhythmic structure, instrumentation, pitch content) with members of other genres [3, 4].

The basis of any system that performs automatic audio analysis is the extraction of feature vectors from the audio pieces [5]. These vectors are sets of numerical data mapping the significant audio features (e.g., tempo, rhythm, pitch, timbre, etc.) to arithmetic values. A vast amount of research has been conducted laying emphasis on different audio features, proposing different feature vector extraction methods and finally different Audio Music Genre Classifications (AMGC) [5-15].

Despite their diversity concerning audio feature vectors for representing an audio piece, most of the proposed AMGC methods train (with these vectors) an Artificial Neural Network (ANN) to achieve high accuracy classification. Although different classifiers (training methods and transfer functions) have already been examined and produced results [1, 5, 13, 16] a very important factor which affects the efficiency of an ANN has been, to best of our knowledge, ignored; ANN architecture.

In this work, the impact of an ANN architecture (i.e., the number of its layers and the number of the neurons in each layer) in its classification results is investigated. Three Traditional Optimization Methods (TOM) and one Evolutionary Computation (EC) in the form of a Genetic Algorithm (GA) are used to optimise the architecture of an ANN performing AMGC.

[+] Corresponding author. Tel.: + 30 2104822222(221); fax: +30 2104821850.
E-mail address: obe@ist.edu.gr.

The rest of the paper is organized as follows. In Section 2 the approach taken is discussed. More specific, the optimization methods are described and how they are incorporate the ANN is explained. The experiments contacted and the results achieved are given in Section 3. The work concludes with some remarks and hints for further research.

2. The Approach

The aim of this work is to build a backpropagation ANN architecture exhibiting high classification accuracy when it classifies music audio pieces into their appropriate genre. Four different techniques are compared. Each of these methods considers the ANN as its fitness function and tries to optimise the architecture of an ANN leading to an increase value of its classification results. This, in fact, means that each optimization method tries to produce the best combination of layers and neurons in each layer so as to minimize classification error of the ANN.

The first approach (called A) is a multidimensional unconstrained nonlinear minimization method, which uses the Nelder-Mead simplex direct search algorithm. It is a direct search method that does not use numerical or analytic gradients and minimizes a function of several variables [17]. The second method used (called B) is a constrained nonlinear multivariable optimization method that uses the large-scale or medium-scale Quasi-Newton algorithm, while the third traditional method (called C) is its unconstrained variation [18, 19]. The fourth method (called D) is a genetic algorithm solver minimizing a given function [20].

All these methods minimize the same ANN's classification error, so as for their results to be comparable. The ANN constructed according to the architecture proposed by each method is trained and tested using data from a real-world music audio collection.

This audio data set is proposed by Tzanetakis and Cook [5] and has been used in several of the aforementioned works. The set consists of 1000 music audio pieces of 30 seconds length each. These pieces are uniformly distributed to 10 music genres: Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae, and Rock. Thus, there are a 100 pieces belonging to each of the 10 genres. Yaslan and Cataltepe [1] used the MARSYAS software [13] in order to extract the audio features from the data set.

In this work, the above feature vectors has been normalized (in [-1,1]) and an extra column has been inserted determining the music genre of each audio piece using integer values from 1 to 10 in order for their data to be used as input and target data for the ANN.

3. Experiments and Results

The optimization methods and the backpropagation ANN are implemented using MATLAB ver. 7.0 of The MathWorks Inc [21]. This widely accepted and powerful mathematic tool provides a variety of toolboxes with built-in algorithms and functions for optimization, neural networks and genetic algorithms implementation and thus it was a natural choice.

During the first set of experiments, in which the number of its layers and the number of neurons in each layer varied, became apparent that the number of layers does not impose any significant variations on the results. Thus, it was decided that the proposed architecture for the ANN should consist of 2 at the most hidden layers (with as many neurons as provided by the optimization methods) and 1 output layer consisting of a single neuron. The hidden layers use the hyperbolic tangent sigmoid transfer function which compiles with the ANN's input data, while the output layer uses the linear transfer function, in order to be able to produce ten (10) different discernible values, as many as the different audio genres that exist in the dataset used.

After the tuning of the experiment each method run (actually, a large number of such runs were conducted in order for statistical significant results to be gained) and the proposed by each method architecture of the ANN built. These ANN were then trained and tested using the audio data set.

The results concerning the proposed architecture by each optimization method is given in Table 1.

Table 1: Neuron consisting each layer (L1, L2)

Methods	L1	L2
A	16	10
B (case 1)	15	5
B (case 2)	7	19
C	15	10
D	11	16

All methods provide quite different results concerning the number of neurons should be used to build the first and the second layer (L1 and L2) of the ANN. Moreover, method B provides two pairs of solutions; equally satisfactory thus reported and used (depicted in Table 1 as case 1 and case 2).

Based on the results, five ANNs are constructed in order to evaluate their classification accuracy results. These ANN are trained and tested using the same dataset and then evaluated with the same "unknown data" set (audio pieces not used in the training and testing of the ANN). The Mean Magnitude Relative Error (MMRE) of each ANN and its classification accuracy results are presented in Table 2.

Table 2: MMRE and Classification Accuracy Results

Neural Network produced by method:	MMRE	Classification Accuracy (%)
A	0.1975	80.25
B (case 1)	0.2068	79.32
B (case 2)	0.1982	80.18
C	0.1902	80.98
D	0.1901	80.99

As it shown, method D (i.e., the one based on the genetic algorithm) proposes the architecture achieving the best classification accuracy for the data set given. Nevertheless, it should be mentioned that the difference between methods C and D is only 0.0001 (i.e., for the given dataset method D is able to classify more precisely only 1 audio piece more than Method C, for every 10000 pieces). Almost the same conclusion applies for all the other combinations. This also became apparent when the ANN's used to classify a set of 20 unknown (applied for the first time) audio pieces. All of them manage to classify 16 out of 20 audio pieces.

4. Conclusions

In this work, the impact of an ANN's architecture in its classification results when it performs automatic audio music genre classification is investigated. Four different optimization techniques are tested in order for the best combination of layers and neurons in each layer to be produced. Using their results 5 different neural networks are built, trained, tested and evaluated using a well-known and broad accepted dataset. The classification accuracy results shown that there are no worth-mentioned differences produced taking into account that no other optimizations, except for the architecture, have been applied to the neural network. It should be mentioned though, that the genetic algorithm results to the best optimization for the given data set. Furthermore, note that the results achieved herein are slightly better than those reported by the previous works mentioned in Section 1. More experimentation using different optimization techniques applied not only on the architecture but also on other factor affecting the performance (e.g., the threshold functions) as well as different types of neural networks and different ways to extract the feature vectors are under investigation.

5. Acknowledgements

We would like to thank Prof. Y. Yaslan and Prof. Z. Cataltepe for providing us with their set of extracted audio features. We would also like to thank Prof. G. Tzanetakis for his suggestions and clarifications.

6. References

[1] Y. Yaslan and Z. Cataltepe. Audio Music Genre Classification Using Different Classifiers and Feature Selection

Methods. *IEEE 18th International Conference on Pattern Recognition* (ICPR'06), 2006.

[2] S. Esmaili, S. Krishnan, K. Raahemifar. Content Based Audio Classification and Retrieval Using Joint Time-Frequency Analysis. *In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP '04). 2004.

[3] Z. Cataltepe, Y.Yaslan, A. Sonmez. Music Genre Classification Using MIDI and Audio Features. In: EURASIP *Journal on Advances in Signal Processing. Hindawi Publishing Corporation.* 2007, Volume 2007, Article ID 36409.

[4] S. Lippens, J. P. Martens, T. De Mulder. A comparison of human and automatic musical genre classification. *In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP '04). May 2004, Montreal, Quebec, Canada, vol. 4, pp. 233–236.

[5] G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing.* 2002, vol.10, no.5, pp. 293-302.

[6] S. Davis and P. Mermelstein. Experiments in Syllable-Based Recognition of Continuous Speech. *IEEE Trans. Acoust., Speech, Signal Processing.* Aug. 1980. vol. 28, pp. 357–366.

[7] J. Foote and S. Uchihashi. The beat spectrum: A new approach to rhythmic analysis. *In Proc. Int. Conf. Multimedia Expo.* 2001.

[8] M. Goto and Y. Muraoka. Music understanding at the beat level: Real-time beat tracking of audio signals. In: Rosenthal D. and Okuno H. (eds.). *Computational Auditory Scene Analysis. Eds. Mahwah, NJ: Lawrence Erlbaum.* 1998, pp. 157–176.

[9] J. Laroche. Estimating tempo, swing and beat locations in audio recordings. *In: Proc. Int. Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA. 2001,* Mohonk, NY, pp. 135–139.

[10] L. Rabiner and B. H. Juang. Fundamentals of Speech Recognition. *Englewood Cliffs, NJ: Prentice-Hall,* 1993.

[11] E. Scheirer. Tempo and beat analysis of acoustic musical signals. J. Acoust. Soc. Amer. Jan. 1998, vol. 103, no. 1, p. 588, 601.

[12] J. Seppänen. Quantum grid analysis of musical signals. *In: Proc. Int. Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).* 2001, Mohonk, NY, pp. 131–135.

[13] G. Tzanetakis and P. Cook. Marsyas: A Software Framework for Audio Analysis. *Organized Sound*. 2000, vol.4, issue 3.

[14] G. Tzanetakis, G. Essl, P. Cook. Automatic musical genre classification of audio signals. *In: Proc. Int. Symp. Music Information Retrieval (ISMIR),* Oct. 2001.

[15] E. Wold, T. Blum, D. Keislar, J. Wheaton. Content-based classification, search, and retrieval of audio. *IEEE Multimedia*. 1996, vol. 3, no. 2.

[16] K. Gurney. An Introduction To Neural Networks. *Taylor & Francis*, 1997.

[17] J.C. Lagarias, J. A. Reeds, M. H. Wright, P. E. Wright. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization.* 1998, Vol. 9, Number 1, pp. 112-147.

[18] T.F. Coleman, and Y. Li. An Interior, *Trust Region Approach for Nonlinear Minimization Subject to Bounds. SIAM Journal on Optimization*. 1996, Vol. 6, pp. 418-445.

[19] T.F. Coleman, and Y. Li. On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds. *Mathematical Programming.* 1994, Vol. 67, Number 2, pp. 189-224.

[20] D. E. Goldberg. Genetic Algorithms in Search, Optimzation & Machine Learning. *Addison-Wesley,* 1989.

[21] MATLAB. The Language of Technical Computing, *The MathWorks Inc., Natick, MA,* 1997.