Estimating Similarity of Musical Rhythm Patterns through the use of a Neural Network Model

Athanasios Fouloulis[1a], Georgios Papadelis[2a], Konstantinos Pastiadis[2b] & George Papanikolaou[1b]

[1a,b] *Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece, E-Mails: [1a] thanos.fouloulis@gmail.com & [1b] pap@eng.auth.gr*

[2a,b] *Department of Music Studies, School of Fine Arts, Aristotle University of Thessaloniki, Greece, E-Mails: [2a] papadeli@mus.auth.gr & [2b] pastiadi@mus.auth.gr*

Abstract

In this work we present the architecture of a system containing two artificial neural networks in cascade - a self-organizing neural network (called SARDNET) and a Multi-Layer Perceptron - that receives a sequence of temporal intervals (performed rhythm pattern) as input and maps it into a given set of prototypical rhythm patterns.

Results provide strong evidence that this type of network architecture may be proven successful on calculating "similarity measures" between a prototypical rhythm pattern and its micro-variations that are consistent with ratings provided by human listeners.

Introduction

Expressive timing is a qualitative aspect of performed music, which refers to all micro-timing deviations of the actual durations of the notes within a rhythmic sequence from their corresponding notational values within western musical tradition (e.g. quarters, quavers, semiquavers e.t.c.) [1].

In the case of short rhythm patterns research has shown that all possible micro-variations obtained when any specific written pattern is performed, are perceptually organized in discrete rhythm (pattern) categories [2]. Each of those categories is represented by a prototype –usually a performed pattern temporally identical to the notated version-. The "goodness" of category membership for any pattern rhythmically similar to the prototype is usually measured either by appropriately selected mathematical "distance metrics" or, alternatively from ratings of rhythmic similarity provided by experienced musicians through pair wise comparison tasks.

Assessment of musical aptitude in young children usually includes the task of rhythm copying, where the child is asked to tap a short rhythm pattern after listening to it. This study examines the potential of an artificial neural network system in providing valid estimates of children's performance accuracy on this task based on a calculation of similarity between any stimulus pattern and its corresponding temporal pattern that is tapped by the child. The system's output for each performed rhythm pattern is then compared with a corresponding goodness-rating value of category membership for this pattern provided by an experienced musician.

The system architecture

The system consists of two artificial neural networks in cascade so that the output of the first -called SARDNET- becomes the input of the second that is a Feedforward Neural Network (Multi-layer Perceptron).

The first network (SARDNET) is a self-organizing neural network that has proven successful for sequence classification problems such as mapping arbitrary sequences of binary and real numbers, phonemic representations of English words [3] and musical rhythm patterns [4]. In this work SARDNET receives as input a sequence of Inter-Onset-Time intervals (IOIs) that correspond to a performed rhythm pattern and generates an activation pattern. This pattern becomes the input of the later network, the Multi Layer Perceptron (Feedforward Neural Network), which maps the performed rhythm pattern into a given set of prototypical rhythm patterns. Each of the neurons that form the network's output layer represents a prototypical rhythm pattern while its activation amplitude reflects a calculated by the network similarity index.

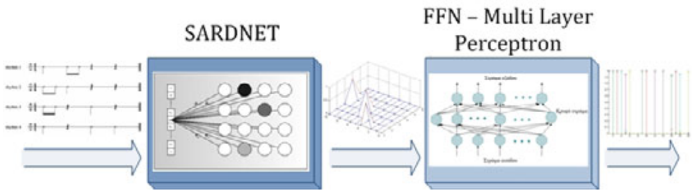

Figure 1: The architecture of the system that consists of two artificial networks in cascade.

The aim of the model proposed here is to generate valid "similarity measures" between prototypical rhythm patterns and their performed variations that are consistent with estimations of rhythmic similarity provided by experienced musicians through listening tests.

Experiments

Our primary concern was to explore the system's potential in generating valid similarity indices between a given set of 20 prototypical rhythm patterns and their micro-variations. Each rhythm pattern was described by a sequence of inter onset intervals (IOIs) between its successive constituent taps. The training data set consisted of 900 rhythm patterns: 300 rhythm patterns that derived from rhythm copying tasks performed by a group of 3rd grade elementary school children, and 600 rhythm patterns that were generated using Gaussian distributions with σ= 10, 20, 30, 40, 50. Furthermore, one experienced musician provided similarity ratings between each prototypical rhythm pattern and their performed micro-variations in a scale from 1 to 10 through a pair wise comparison task.

First we trained SARDNET with the data set of these 900 rhythm patterns. We used a map of 8x8 elements and a learning rate of 0.08. The network was trained for 100 epochs and after the completion of self – organization the activation maps for all the elements (900) of the training data set were utilized to train the Multilayer Perceptron (MLP).

The Feedforward network (MLP) which was used for this experiment had 64 (8x8) input units, that received the activation values directly from the output layer of the SARDNET network. These input units were fully connected to a hidden layer with 24 neurons, which were also fully connected to 20 output neurons. Each of them corresponded to a prototypical rhythm pattern of the stimulus rhythm set.

The MLP was trained using the 900 activation patterns (maps) from the SARDNET and their corresponding similarity indices that the experienced musician provided as target values. This process reached an error of $2{,}8 \times 10^{-4}$ in about 10000 epochs.

The MLP was trained in such a way that if any of the prototypical rhythm patterns was presented in its input, the system should activate the corresponding neuron from the output layer with maximum amplitude. Besides, if a performed rhythm pattern was presented the system should activate an output neuron that corresponds to the category to which that pattern would most-likely belong, with an amplitude that should be equivalent to the similarity rating for that pattern given by a musically experienced person.

Evaluating network's efficiency

In order to evaluate the potential of the system in generating valid indices of similarity between the prototypical patterns and their variations we used a second data set. This data set contained 220 rhythm patterns that were obtained through rhythm copying tasks from a group of 2nd grade elementary school children. Similarity ratings for all patterns of this data set were also provided by the same experienced musician.

The correlation coefficient between the ratings provided by the experienced musician and those that were calculated from the network was 0.71. However, 20% of the test data set was excluded from this calculation because the system was not successful identifying the "correct" pattern category, thus activating the right output neuron. After examining the data we found out that this behavior occurred for two different reasons: firstly, it was observed that in cases where the performed rhythm patterns had quite big differences compared to their respective prototypes and, secondly, in cases where the rhythm pattern performed by the child was obviously similar to another prototypical pattern of the stimulus data set. As for the first condition the cause of the system's failure might be due to the fact that it was trained only with the specific data set of rhythm patterns. Consequently if any of the performed rhythm patterns were far different from their respective prototypes the system was unable to recognize it correctly and therefore to activate the corresponding output neuron with appropriate amplitude. In the second case the system's failure shouldn't constitute a problem per se since the presented input rhythm activated the most appropriate neuron of the output layer.

Discussion

The architecture presented above still has many open issues that need to be explored. First of all the system's performance should be examined considering a larger training data set which will contain also variations that are far different from their respective prototypical patterns. This way the system's tolerance in processing different categories of performance accuracy on the rhythm copying task will be defined.

Furthermore the system should be trained to provide a set of "similarity measures" for each performed rhythm presented in its input. Each item of that set would refer to each one of the prototypical patterns of the entire stimulus rhythm set. In order to achieve this it is necessary to train the MLP network with targets that contain respective similarity values to the entire set of prototypical patterns for each performed rhythm pattern. Therefore, such an informationally richer data set of ratings should be collected from musically experienced listeners.

Last, it is clear that system's training should also be based on similarity rating data drown from a larger population of listeners, so that it would not reflect only an individual's judgments.

Conclusion

In this work we presented the architecture of a system containing two artificial neural networks in cascade that receive a sequence of temporal intervals as input and maps it into a given set of stimulus rhythm patterns. Results provide strong evidence that this type of network architecture may be proven successful on calculating "similarity measures" between a rhythm pattern and its micro-variations that are consistent with ratings provided by musically experienced listeners.

References

[1] Desain, P & Honning, H.: Music, Mind and Machine. Thesis Bublishers, Amsterdam, 1992

[2] Papadelis, G. & Papanikolaou, G.: The Perceptual Space Between and Within Musical Rhythm Categories. In, J. Davidson (ed.): The Music Practitioner. Ashgate, Aldershot, 2004

[3] Miikkulainen, R. & James, D. L.: SARDNET: A Self-Organizing Feature Map for Sequences. In G. Tesauro, D. S. Touretzky, and T. K. Leen (eds): Proceedings of the Advances in Neural Information Processing Systems, vol. 7, Morgan Kaufmann, 1995

[4] Martins, J. M., Gimenes, M., Manzolli, J. & Maia, Jr. A.: Similarity Measures for Rhythmic Sequences. Proceedings of the 10th Brazilian Symposium of Musical Computation, Belo Horizonte (Brazil), 2005