25th Italian Workshop on Neural Networks (Vietri sul Mare)

Benchmarking Functional Link Expansions for Audio Classification Tasks

Scardapane S., Comminiello D., Scarpiniti M.,
Parisi R. and Uncini A.

Overview

Audio classification

Audio classification is the task of automatically assigning one or more labels to a song. It includes the following tasks:

- Genre classification (e.g. Rock/Pop),
- Author recognition,
- Perception of mood,
- Speech/music discrimination,
- Leading instrument identification,
- etc.

Audio classification is a fundamental component for any music information retrieval (MIR) system [FLTZ11].

Machine learning and audio classification

From a machine learning perspective, audio classification consists of the following aspects:

Song representation The input to a classifier is a set of d features extracted from a song. This is the *representation* problem.

Training set collection This is the task of collecting and correctly labeling an initial training set of songs.

Choice of classifier Choosing a suitable classifier (and its hyper parameters) is essential for optimal performance.

In this paper we focus on the third aspect, on a particular class of neural networks known as functional link NNs (FLNNs).

Functional Link NNs

A FLNN processes its input with two successive operations [Pao89]:

1. A *fixed* nonlinear expansion via a *functional link* expansion block.
2. A trainable linear filtering operation.

FLNNs have been successfully applied to the task of audio classification [SCSU13], but one major question remains open:

How to suitably choose the proper functional expansion block?

In this work, we aim at providing some guidelines to this question in the case of audio classification.

Overview

Architecture of a FLNN

Given an input vector $\mathbf{x} \in \mathbb{R}^d$, the output of an FL network is computed as:

$$f(\mathbf{x}) = \sum_{i=1}^{B} \beta_i h_i(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{h}(\mathbf{x}) \,, \tag{1}$$

where each $h_i(\cdot)$ is a fixed non-linear term, denoted as functional-link. The overall vector:

$$\mathbf{h}(\cdot) = [h_1(\cdots), \ldots, h_B(\cdot)]^T \tag{2}$$

is called the *functional expansion* block.

Least-square training

Given a dataset of N pairs song/class for training, denoted as $T = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_L, y_N)\}$, let:

$$\mathbf{H} = \begin{bmatrix} h_1(\mathbf{x}_1) & \cdots & h_B(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_N) & \cdots & h_B(\mathbf{x}_N) \end{bmatrix} \tag{3}$$

and $\mathbf{y} = [y_1, y_2, \ldots, y_N]^{\mathrm{T}}$ be the hidden matrix and the output vector respectively. The optimal weights are obtained by solving:

$$\min_{\boldsymbol{\beta}} \frac{1}{2} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{y}\|^2 + \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2, \tag{4}$$

where $\lambda > 0$ is a regularization factor.

Chebyshev polynomial expansion

The Chebyshev polynomial expansion for a single feature x_j of the pattern $\mathbf{x}$ is computed recursively as:

$$h_k\left(x_j\right) = 2x_j h_{k-1}\left(x_j\right) - h_{k-2}\left(x_j\right) \ , \tag{5}$$

for $k = 0, \ldots, P-1$, where P is the *expansion order*. The overall expansion block is then obtained by concatenating the expansions for each element of the input vector. In (5), initial values (i.e., for $k = 0$) are:

$$\begin{aligned} h_{-1}\left(x_j\right) &= x_j \ , \\ h_{-2}\left(x_j\right) &= 1. \end{aligned} \tag{6}$$

Legendre polynomial expansion

The Legendre polynomial expansion is defined for a single feature x_j as:

$$h_k\left(x_j\right) = \frac{1}{k}\left\{\left(2k-1\right)x_j h_{k-1}\left(x_j\right) - \left(k-1\right)h_{k-2}\left(x_j\right)\right\} \tag{7}$$

for $k = 0, \ldots, P-1$. Initial values in Eq. (7) are set as before.

Trigonometric series expansion

The trigonometric basis expansion is given by:

$$h_k(x_j) = \begin{cases} \sin\left(p\pi x_j\right), & k = 2p-2 \\ \cos\left(p\pi x_j\right), & k = 2p-1 \end{cases}, \tag{8}$$

where $k = 0, \ldots, B$ is the functional link index and $p = 1, \ldots, P$ is the expansion index, being P the expansion order. Cross-products between elements of the pattern $\mathbf{x}$ can also be considered.

Random vector expansion

The random vector (RV) expansion is parametric with respect to a set of internal weights, that are stochastically assigned. A RV functional link (with sigmoid nonlinearity) is given by:

$$h_k(\mathbf{x}) = \frac{1}{1 + e^{(-\mathbf{a}\mathbf{x}+b)}}\,, \tag{9}$$

where the parameters $\mathbf{a}$ and b are randomly assigned at the beginning of the learning process. Unlike the previous expansion types, the overall number B of functional links is a free parameter in this case, while in the previous expansions it depends on the expansion order.

Overview

Experimental Setup

Table : General Description of The Datasets.

Dataset name	Features	Instances	Task	Classes	Reference
Garageband	49	1856	Genre recognition	9	[MM05]
Artist20	30	1413	Artist recognition	20	[Ell07]
GTZAN	13	120	Speech/Music Discrimination	2	[TC02]

- We perform a 3-fold cross-validation on the available data, repeated 10 times.
- We optimize the models by performing a grid search procedure, using an inner 3-fold cross-validation on the training data.
- In all cases, input features were normalized between -1 and $+1$ before the experiments.

Experimental results

Table : Final misclassification error and training time for the four functional expansions, together with standard deviation. Best results in boldface.

Dataset	**Algorithm**	**Error**	**Time** [secs]
Garageband	TRI-FL	0.415 ± 0.013	0.156 ± 0.030
	CHEB-FL	0.407 ± 0.0126	$\mathbf{0.055 \pm 0.001}$
	LEG-FL	$\mathbf{0.404 \pm 0.0140}$	0.072 ± 0.026
	RV-FL	0.411 ± 0.017	0.090 ± 0.015
Artist20	TRI-FL	0.410 ± 0.016	0.084 ± 0.013
	CHEB-FL	0.401 ± 0.021	$\mathbf{0.037 \pm 0.001}$
	LEG-FL	0.442 ± 0.020	0.040 ± 0.001
	RV-FL	$\mathbf{0.375 \pm 0.018}$	0.070 ± 0.014
GTZAN	TRI-FL	0.316 ± 0.073	$\mathbf{0.001 \pm 0.002}$
	CHEB-FL	0.317 ± 0.066	0.004 ± 0.001
	LEG-FL	0.334 ± 0.071	0.004 ± 0.001
	RV-FL	$\mathbf{0.222 \pm 0.062}$	0.005 ± 0.002

ROC curve for GTZAN

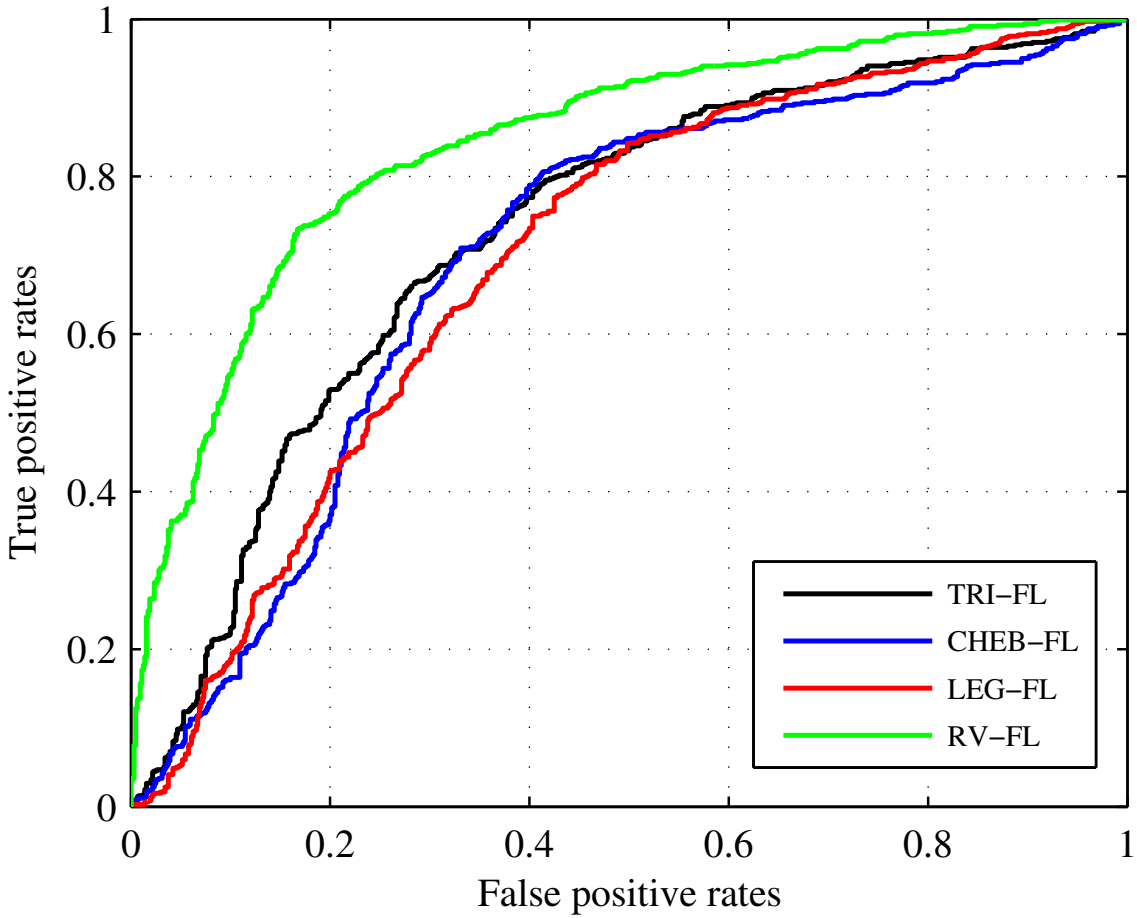

Figure : ROC curve for the GTZAN dataset.

Overview

Conclusions

- FLNN are efficient models for audio classification tasks, but their performance strongly depends on the functional expansion block.
- We presented an analysis of several expansions, considering three different tasks, including genre and artist recognition.
- Our experimental results suggest that the random vector expansion outperforms other common choices, while requiring a comparable training time.

References

D. P. W. Ellis.
Classifying music audio with timbral and chroma features.
In *Proceedings of the 8th International Conference on Music Information Retrieval*, pages 339–340. Austrian Computer Society, 2007.

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.

I. Mierswa and K. Morik.
Automatic feature extraction for classifying audio data.
Machine learning, 58(2-3):127–149, 2005.

Y.-H. Pao.
Adaptive Pattern Recognition and Neural Networks.
Addison-Wesley, Reading, MA, 1989.

S. Scardapane, D. Comminiello, M. Scarpiniti, and A. Uncini.
Music Classification Using Extreme Learning Machines.
In *8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 377–381, Trieste, Italy, September 2013.

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