

A Wrapper for Projection Pursuit Learning

Leonardo M. Holschuh, Clodaldo A. M. Lima, Fernando J. Von Zuben (Orientador)

Departamento de Engenharia de Computação e Automação Industrial (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (Unicamp)
Caixa Postal 6101, CEP 13083-970 – Campinas, SP, Brasil

{holschuh, moraes, vonzuben}@dca.fee.unicamp.br

Abstract – Constructive algorithms are effective methods for designing Artificial Neural Networks (ANN) with good accuracy and generalization capability, yet with parsimonious network structures. Projection Pursuit Learning (PPL) has demonstrated great flexibility and effectiveness in performing this task, though presenting some difficulties in the search for appropriate projection directions in input spaces with high dimensionality. This characteristic hinders the use of PPL at the field of time series prediction, due to the common occurrence of high-dimensional input spaces. We propose a method based on the wrapper methodology to perform variable selection, so that only a subset of highly informative lags is going to be considered as the regression vector. The results show that variable selection increases the performance of the final ANN at an acceptable increase of computational cost.

Keywords – Artificial Neural Networks, Constructive Algorithms, Projection Pursuit Learning, Variable Selection, Wrapper Approach, Time-Series Prediction.

1. Introduction

The generalization capability of an Artificial Neural Networks (ANNs) [3] is highly dependent on its size, among other design aspects. With too few hidden neuron, it might not be possible to approximate the desired input-output behavior. With too many, we have the overfitting problem. Though previous experiences may in some cases guide the ANNs designer toward appropriate architectures, this process is still basically, guesswork.

To avoid the “dangers” of guessing architectures, the usual solution is to use one of the several methods of automatic architecture determination available in the literature.

Constructive methods have a higher bias for topologies with less neurons than other architecture determination methods, and amongst the constructive methods, the Projection Pursuit Learning (PPL) algorithm [5], [6] has presented great flexibility and effectiveness.

The PPL algorithm is capable of significantly reducing data dimensionality. The role of each neuron in the approximation of the desired input-output mapping is defined during the learning phase, including the estimation of the most suitable activation function. At the end, the constructive algorithm will guide to an one-hidden layer architecture similar to a Multilayer Perceptron (MLP) topology, except for the existence of a specific activation function for each neuron in the hidden layer.

However, the PPL algorithm may have its efficiency noticeably affected by the presence of low-relevant variables in the training dataset, and the overall performance may be significantly improved when a variable selection procedure [2] is applied in conjunction with PPL. We will adopt here a wrapper approach [2] as it permits us to search for a subset of input variables that is optimal in relation to PPL.

2. Proposal

As our purpose is to obtain parsimonious ANNs presenting high generalization capability using the PPL algorithm in high-dimensional input-space problems, the wrapper methodology was chosen to guide the search for an optimal set of input variables. The search will be performed by a standard genetic algorithm (GA) [1], using chromosomes with simple binary representation (one bit for each candidate input, ones indicating that the corresponding inputs have been selected, zeros indicating that the corresponding inputs have not been selected).

Time-series prediction was chosen as the primary class of problems to test the methodology, and in this context, candidate inputs are usually the previous values of the series. If, for example, we wish to predict the value of a series at the time instant t , we might take the value of the series at the time instants $t-1, t-2, \dots, t-p$ as inputs to the ANN, where p is the size of the lag window.

Fitness is determined by the normalized mean squared error (NMSE) of the ANN, trained

using the subset of inputs determined by the binary vector over a validation dataset. The other particulars of the GA are: single-point crossover, complementary mutation, roulette-wheel selection, fixed population size, and elitism (keeping both the best and worst chromosomes of the previous generation).

We have called this methodology *Wrapper for Input Selection in Projection Pursuit Learning* (WIS-PPL for short). Its pseudocode is given bellow:

```

Initialize first generation of candidate subsets of inputs,
For Gen = 1 to MaxGen,
  For Subset = 1 to last,
    Train ANN using PPL and Subset,
    Calculate Fitness based on the validation error,
  End For
  Determining next Gen:
    Copy the Subsets with the best and the worst
    Fitnesses to the next Generation,
  For remaining Subset slots,
    Roll roulette wheel to choose pair,
    Check for crossover,
    Check for mutation,
  End For
End For

```

3. Results

For reasons of space, we present in the following tables results obtained for the yearly sunspot number time series only. The maximum lag window considered was of 11 years (the series is cyclic with an 11 years period).

Table 1 compares the WIS-PPL algorithm, a single run of PPL using all 11 inputs, and several other algorithms listed bellow. The training and validation data-sets were selected randomly from the year of 1900 to 1920; and the remaining instances were separated in three different test sets: Test1 (1921-1955), Test2 (1956-1979), and Test3 (1980-1994).

The other algorithms listed [4] are: Auto-Regressive models (AR), Weight Elimination Feed Forward Network (Wnet), Dynamical Recurrent Neural Network (DRNN), Soft Weight Sharing Network (SSNet), Scale Neural Network (ScaleNet), Wan's Committee Prediction method (COMM), recurrent networks trained with the Backpropagation Through Time algorithm (BPTT), Constructive BBTT (CBPTT), Violation Guided Back Propagation (VGBP), and the Marra and Morabito's technique based on "de-rectification" of the sunspot data and Elman Networks (DDEN).

4. Conclusions

The results indicate that a wrapper for projection pursuit learning is competitive, at least when applied to the synthesis of time series predictors

with a moderate lag window.

The learning time for the WIS-PPL algorithm is considerably high, as was expected for a wrapper-based method, but the possibility of obtaining a parsimonious prediction model, and better results than the one obtained without variable selection, points favorably to the use of the computational resources needed.

Design Method	NMSE		
	Test1	Test2	Test3
AR(12)	0.427	0.966	0.238
Wnet	0.086	0.350	0.219
SSNet	0.077	N/A	N/A
DRNN	0.091	0.273	N/A
COMM	0.065	0.240	0.148
ScaleNet	0.057	0.130	N/A
BPTT	0.084	0.300	N/A
CBPTT	0.092	0.251	N/A
VGBP	0.033	0.052	0.033
DDEN	0.043	0.080	0.028
PPL, 11 inputs	0.073	0.048	0.063
WIS-PPL, mean	0.053	0.067	0.055
WIS-PPL, best	0.046	0.060	0.035

WIS-PPL, mean refers to the mean result of 10 different runs of the algorithm, WIS-PPL, best is the best of these runs..

Table 1 - Comparison of WIS-PPL and Other Design Methods

References

- [1] T. Bäck, D. B. Fogel, and Z. Michalewicz (editors), *Evolutionary Computation 1 and 2*. IOP Publishing, 2000.
- [2] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
- [3] K. Hornik, M. Stinchcombe, H. White and P. Auer, "Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives," *Neural Computation*, vol. 6, no. 6, pp. 1262-1275, 1994.
- [4] S. Marra and F. C. Morabito, "A New technique for solar activity forecasting using recurrent Elman networks," *International Journal of Computational Intelligence*, Vol. 3, no. 1, 2006.
- [5] F. J. Von Zuben and M. L. A. Netto, "Unit-growing learning optimizing the solvability condition for model-free regression," *Proc. IEEE International Conference on Neural Networks*, pp. 795-800, 1995.
- [6] F. J. Von Zuben, and M. L. A. Netto, "Projection pursuit and the solvability condition applied to constructive learning," *Proc. IEEE International Conference on Neural Networks*, pp. 1062-1067, 1997.