ABSTRACT

A time series is a sequence of real values that can be considered as observations from a certain system. In this work, we are interested in time series coming from dynamical systems. Such systems can be sometimes described by a set of equations that model the underlying mechanism from where the samples come. However, in several real systems, those equations are unknown, and the only information available is a set of temporal measures, that constitute a time series. On the other hand, by practical reasons it is usually required to have a prediction, v.g. to know the (approximated) value of the series in a future instant *t*. The goal of this thesis is to solve one of such real-world prediction problem: given historical data related with the liquefied bottled propane gas sales, predict the future gas sales, as accurately as possible. This time series prediction problem is addressed by means of neural networks, using both (dynamic) reconstruction and prediction. The problem of dynamically reconstruct the original system consists in building a model that captures certain characteristics from the original system in order to have a correspondence between the model and the long-term behavior of the system.

The network design process is basically guided by three ingredients. The dimensionality of the problem is explored by our first ingredient, **Takens-Mañé's theorem**. By means of this theorem, the optimal dimension of the (neural) network input can be investigated. Our second ingredient is a strong theorem: neural networks with a single hidden layer are universal approximators. As the third ingredient, we faced the search of the optimal size of the hidden layer by means of genetic algorithms, used to suggest the number of hidden neurons that maximizes a target fitness function (related with prediction errors). These algorithms were also used to find the most influential networks inputs in some cases. The determination of the hidden layer size is a core (and hard) problem in the determination of the network topology. As was mentioned, we explored also.

This thesis includes a state of the art of neural networks design for time series prediction, including related topics such as dynamical systems, universal approximators, gradient-descent searches and variations, as well as meta-heuristics. Special attention is given to the available software tools for neural networks design and time series processing. After a review of the available software packages, the most promising computational tools for both approaches are discussed, and a state of the art in the theory of neural networks for the prediction of time series is also included. The criteria of network selection are discussed and a trade-off between performance and network complexity is further explored, inspired in a minimum description length and its estimation. Network committees showed to be economical solutions, where the predictions are a naive convex combination of predictions from individual networks.

The need of criteria to compare the behavior of the model with the real system, in the long run, for a dynamic stochastic systems, is presented and two alternatives are commented.

The obtained results proof the existence of a solution to the problem of learning of the dependence **Input** → **Output**. This solution was found in a constructive and exhaustive way. The exhaustiveness can be deduced from the next five statements:

1) the design of a neural network requires knowing the input and output dimension, the number of the hidden layers and of the neurons in each of them.
2) the use of the Takens-Mañé's theorem allows to derive the dimension of the input data
3) by theorems such as the Kolmogorov's and Cybenko's ones the use of MLPs with only one hidden layer is justified so several of such models were tested, because of their simplicity
4) the number of neurons in the hidden layer is determined many times heuristically using genetic algorithms
5) a neuron in the output gives the desired prediction

Finally, the results confirmed several already known facts (such as that adding noise to the inputs and outputs of the training values can improve the results; that recurrent networks trained with the back-propagation algorithm don't have the problem of vanishing gradients in short periods and that the use of committees - which can be seen as a very basic of distributed artificial intelligence - allows to improve significantly the predictions).