University of Salzburg

Institute of Computer Science

SatMAS - Input pattern selection in an MAS

Thomas Aschauer

Thorsten Engl

Date: March 21, 2001

Version: 1.0

Contents

List of Figures

Chapter 1

Introduction

This chapter gives an overview about motivation for input pattern selection in an MAS. Additionally we give an overview of already existing methods of resolution.

1.1 Motivation

When training an ANN we often face the problem of huge amounts of possible training data. This would result in extreme time consuming training cycles if all available data is used for teaching the network. So it's important to select only the essential part of the training pattern and keep the training data set as small as possible.

On the other hand, minimization of generalization error has also to be guaranteed, in order to ensure that the trained network can properly yield an optimal result. The selection of training data presented to the neural network influences whether or not the network learns a particular task. Like a child, how well a network will learn depends on the examples presented. A good set of examples, which illustrate the tasks to be learned well, is necessary for the desired learning to take place. The set of training examples must also reflect the variability in the patterns that the network will encounter after training. At first glance this appears to be a contradictory pair of objectives. However, just as the generalization error must be as low as possible, data sampling which involves both collection and measurement of data is expensive and therefore needs to be reduced to a minimum. There are several methods of resolution for selecting a well fitting subset of the original (in most cases very huge) data set. However, the problem of selecting the optimal training set has not yet been solved.

1.2 Existing Methods of Input Pattern Selection

- Active Learning
  - Active Selection
  - Active Sampling
- Dynamic Pattern Selection
- Training Data Selection with Genetic Algorithms

- Training Data Selection with Multi Layer Neural Networks

1.2.1 Active Selection

The starting point for active learning is the observation that the traditional approach of randomly selecting training samples leads to large, highly redundant training sets. Such training sets can be obtained if the learner is enabled to select those training data that he/she expects to be most informative. In this case, the learner is no longer a passive recipient of information but takes an active role in the selection of the training data.

A deeper introduction to active learning can be achieved in [1].

1.2.2 Active Sampling

Recent research has shown active learning methods to be effective in increasing the modeling reliability of a neural network system. An active learning agent has the ability to query its environment in order to make a selection of its training data. One approach to the implementation of active leaning is to use querying-by-committee. This results in considerably reduced data collection and at the same time does not compromise the accuracy of identification. A nonlinear plant with both clean and noisy data is successfully modeled by such a technique and a feed forward neural network controller based upon such a model is demonstrated to perform effectively.

Minimized Data Collection - Active Querying Example

This is a data gathering method based on active querying. In this method data is reduced to a minimum, yet modeling accuracy is not compromised. The active querying criterion is determined by whether or not several neural network models agree when they are fitted to random sub samples of a small amount of collected data.

For details see [2].

1.2.3 Dynamic Pattern Selection

In contrast to active pattern selection, the dynamic pattern selection algorithm achieves concise training sets by continually validating the generalization properties of the net.

Details of this method can be found in [3].

1.2.4 Training Data Selection with Genetic Algorithms

In this method a genetic algorithm is employed for the parallel selection of appropriate input pattern for the training data set.

For an example see [4].

1.2.5 Training Data Selection with Multi Layer Neural Networks

This method selects a small number of training data, which guarantee both generalization and fast training of the MLNNs applied to pattern classification. The generalization will be satisfied using the data locate close to the boundary if the pattern classes. However, if these data are only used in the training, convergence is slow. Therefore the MLNN is first trained using some number of the data, which are randomly selected (Step 1). The data, for which the output error is relatively large, are selected. Furthermore, they are paired with the nearest data belong to the different class. The newly selected data are further paired with the nearest data. Finally, pairs of data, which locate close to the boundary, can be found. Using these pairs of the data, the MLNNs are further trained(Step 2). Since, there are some variations to combine Steps 1 and 2, the proposed method can be applied to both off-line and on-line training. The proposed method can reduce the number of the training data, at the same time, can hasten the training.

A detailed description can be found at [5].

Appendix A

Reports

This chapter contains our reports and results of the week.

Bibliography

[1] H. Ritter M. Hasenjäger. Active learning in neural networks. Technical report, Technische Fakultät, Universität Bielefeld, Germany, 1998.

[2] Leonard G.C. Hamey Tirthankar RayChaudhuri. Accurate modelling with minimised data collection – an active learning algorithm. Technical report, School of MPCE, Macquarie University, New South Wales 2109, Australia, 1995.

[3] Axel Röbel. The dynamic pattern selection algorithm: Effective training and controlled generalization of backpropagation neural networks. Technical report, TU Berlin, Germany, 1998.

[4] Helmut Mayer Roland Schwaiger. Is evolution a good teacher? genetic algorithms create training data sets for artificial neural networks. Technical report, Department of Computer Science, University of Salzburg, A-5020 Salzburg, Austria, 1998.

[5] Kenji NAKAYAMA Kazuyuki HARA. Training data selection method for generalization by multilayer neural networks. Technical report, 1998.