IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB8617

Chapter II

A Perturbation Size-Independent Analysis of Robustness in Neural Networks by Randomized Algorithms

C. Alippi
Politecnico di Milano, Italy

ABSTRACT

This chapter presents a general methodology for evaluating the loss in performance of a generic neural network once its weights are affected by perturbations. Since weights represent the "knowledge space" of the neural model, the robustness analysis can be used to study the weights/performance relationship. The perturbation analysis, which is closely related to sensitivity issues, relaxes all assumptions made in the related literature, such as the small perturbation hypothesis, specific requirements on the distribution of perturbations and neural variables, the number of hidden units and a given neural structure. The methodology, based on Randomized Algorithms, allows reformulating the computationally intractable problem of robustness/sensitivity analysis in a probabilistic framework characterised by a polynomial time solution in the accuracy and confidence degrees.

INTRODUCTION

The evaluation of the effects induced by perturbations affecting a neural computation is relevant from the theoretical point of view and in developing an embedded device dedicated to a specific application.

In the first case, the interest is in obtaining a reliable and easy to be generated measure of the performance loss induced by perturbations affecting the weights of a neural network. The relevance of the analysis is obvious since weights characterise the “knowledge space” of the neural model and, hence, its inner nature. In this direction, a study of the evolution of the network’s weights over training time allows for understanding the mechanism behind the generation of the knowledge space. Conversely, the analysis of a specific knowledge space (fixed configuration for weights) provides hints about the relationship between the weights space and the performance function. The latter aspect is of primary interest in recurrent neural networks where even small modifications of the weight values are critical to performance (e.g., think of the stability of an intelligent controller comprising a neural network and issues leading to robust control).

The second case is somehow strictly related to the first one and covers the situation where the neural network must be implemented in a physical device. The optimally trained neural network becomes the “golden unit” to be implemented within a finite precision representation environment as it happens in mission-critical applications and embedded systems. In these applications, behavioural perturbations affecting the weights of a neural network abstract uncertainties associated with the implementation process, such as finite precision representations (e.g., truncation or rounding in a digital hardware, fixed or low resolution floating point representations), fluctuations of the parameters representing the weights in analog solutions (e.g., associated with the production process of a physical component), ageing effects, or more complex and subtle uncertainties in mixed implementations.

The sensitivity/robustness issue has been widely addressed in the neural network community with a particular focus on specific neural topologies.

More in detail, when the neural network is composed of linear units, the analysis is straightforward and the relationship between perturbations and the induced performance loss can be obtained in a closed form (Alippi & Briozzo, 1998). Conversely, when the neural topology is non-linear, which is mostly the case, several authors assume the small perturbation hypothesis or particular hypothesis about the stochastic nature of the neural computation. In both cases, the assumptions make the mathematics more amenable with the positive consequence that a relationship between perturbations and performance loss can be derived (e.g., see Alippi & Briozzo, 1998; Pichè, 1995). Unfortunately, these analyses introduce hypotheses which are not always satisfied in all real applications.