Neural Networks 54 (2014) 17–37

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Solving the linear interval tolerance problem for weight initialization of neural networks

S.P. Adam [a,b,*], D.A. Karras [c], G.D. Magoulas [d], M.N. Vrahatis [a]

[a] Computational Intelligence Laboratory, Department of Mathematics, University of Patras, GR-26110 Patras, Greece

[b] Department of Computer Engineering, Technological Educational Institute of Epirus, 47100 Arta, Greece

[c] Department of Automation, Technological Educational Institute of Sterea Hellas, 34400 Psahna, Evia, Greece

[d] Department of Computer Science and Information Systems, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

ARTICLE INFO

Article history:
Received 30 March 2013
Received in revised form 3 February 2014
Accepted 13 February 2014
Available online 24 February 2014

Keywords:
Neural networks
Weight initialization
Interval analysis
Linear interval tolerance problem

ABSTRACT

Determining good initial conditions for an algorithm used to train a neural network is considered a parameter estimation problem dealing with uncertainty about the initial weights. Interval analysis approaches model uncertainty in parameter estimation problems using intervals and formulating tolerance problems. Solving a tolerance problem is defining lower and upper bounds of the intervals so that the system functionality is guaranteed within predefined limits. The aim of this paper is to show how the problem of determining the initial weight intervals of a neural network can be defined in terms of solving a linear interval tolerance problem. The proposed linear interval tolerance approach copes with uncertainty about the initial weights without any previous knowledge or specific assumptions on the input data as required by approaches such as fuzzy sets or rough sets. The proposed method is tested on a number of well known benchmarks for neural networks trained with the back-propagation family of algorithms. Its efficiency is evaluated with regards to standard performance measures and the results obtained are compared against results of a number of well known and established initialization methods. These results provide credible evidence that the proposed method outperforms classical weight initialization methods.

1. Introduction

The purpose of Interval Analysis (IA) is to set upper and lower bounds on the effect produced on some computed quantity by different types of mathematical computing errors (rounding, approximation, uncertainty etc.) (Hansen & Walster, 2004; Moore, 1966). Intervals are used to model uncertainty in parameter estimation problems such as the noise associated with measured data. Such problems arise in engineering design or mathematical modeling where tolerances in the relevant parameters need to be defined in terms of upper and lower bounds so that the desired functionality is guaranteed within these bounds. The interval-based algorithms are used to reliably approximate the set of consistent values of parameters by inner and outer intervals and thus take into account all possible options in numerical constraint satisfaction problems.

The promising features of IA motivated researchers from different disciplines to invest in the study and implementation of IA methods whenever reliable numerical computations are required. Currently, this research field is rapidly growing due to the increasing computation power of modern hardware. Examples of applications range from finite element analysis (Degrauwe, Lombaert, & Roeck, 2010) and data analysis (Garloff, Idriss, & Smith, 2007), to stock market forecasting (Hu & He, 2007), reliability of mechanical design (Penmetsa & Grandhi, 2002), and many more. Research in the area of neural networks has also benefited from IA and a number of efforts utilizing concepts and methods from IA are reported in the literature. Examples are those by de Weerdt, Chu, and Mulder (2009) on the use of IA for optimizing the neural network output, Ishibuchi and Nii (1998) on the generalization ability of neural networks, Xu, Lam, and Ho (2005) on robust stability criteria for interval neural networks, Li, Li, and Du (2007) regarding training of neural networks, and others.

An important problem encountered when training a neural network is to determine appropriate initial values for the connection weights. Effective weight initialization is associated to

* Corresponding author at: Computational Intelligence Laboratory, Department of Mathematics, University of Patras, GR-26110 Patras, Greece. Tel.: +30 6970806559.

E-mail address: adamsp@upatras.gr (S.P. Adam).

http://dx.doi.org/10.1016/j.neunet.2014.02.006

performance characteristics such as the time needed to successfully train the network and the generalization ability of the trained network. Inappropriate weight initialization is very likely to increase the training time or even to cause non convergence of the training algorithm, while another unfortunate result may be to decrease the network's ability to generalize well, especially when training with back-propagation (BP), a procedure suffering from local minima, (Hassoun, 1995; Haykin, 1999; Lee, Oh, & Kim, 1991). These are defaults and limitations for having successful practical application of neural networks in real life processes.

The importance manifested by the research community for this subject has been demonstrated by the number of research work published in this area. The proposed approaches can be, roughly, divided into two categories. Methods in the first category perform input data clustering in order to extract significant information (feature vectors or reference patterns) pertaining the pattern space and initial connection weights are chosen to be near the centers of these clusters. The main drawback of these methods is the computational cost needed to preprocess the input data. Often this cost may be prohibitive for these methods to be used in real world applications. The second category includes those methods that are based on random selection of initial weights from a subset of $\mathbb{R}^n$, which is an interval defined considering important properties of the pattern space and/or the parameters of the training process.

The notion of the interval, underlying random weight selection methods, suggests the idea to use IA in order to deal with uncertainty about the initial weights. Hence, the unknown initial weights are considered to be intervals with unknown bounds. Under generally adopted assumptions about the input to any node, the resulting unknown interval quantity is then limited within specific upper and lower bounds. Ensuring that scientific computations provide results within guaranteed limits is an issue mentioned by researchers in IA as a tolerance problem. In consequence, the approach proposed herein gives rise to formulating a linear interval tolerance problem which is solved to determine significant intervals for the initial weights. Beaumont and Philippe (2001), Pivkina and Kreinovich (2006), Shary (1995) and other researchers propose different methods for solving a tolerance problem. Besides formulating the problem of determining initial weights as a linear interval tolerance problem, we also present here a new algorithm for defining the required solution to the specific tolerance problem.

The proposed linear interval tolerance approach (LIT-Approach) deals with uncertainty about the initial weights based exclusively on numerical information of the patterns without any assumption on the distribution of the input data. IA provides the means of handling uncertainty in parameters in much the same way this happens with other approaches such as the possibilistic approach with Fuzzy sets (Zadeh, 1978), Evidence theory (Shafer, 1976), Rough sets (Pawlak, 1991) or methods combining properties of these approaches. However, methods using fuzzy sets require parameters of the membership functions to be tuned and eventually some preprocessing of the input data to be done if pertinent input variables need to be identified. Moreover, when using rough sets one needs to process the input data in order to deal with the indiscernibility relation and establish upper and lower approximations of the concepts pertaining the problem, see Bello and Verdegay (2012). Finally, application of the Dempster–Shafer (evidence) theory is a matter of subjective estimation of uncertainty as it assumes that values of belief (or plausibility) are given by an expert. Unlike all these approaches, the interval computation used for LIT-Approach needs only elementary statistics of the input data to be computed such the sample mean, the sample standard deviation or the median and the quartiles of the sample.

It is worth noting here the approach formulated by Jamett and Acuña (2006) as an interval approach for weight initialization. The solution proposed "solves the network weight initialization problem, performing an exhaustive search for minima by means of interval arithmetic. Then, the global minimum is obtained once the search has been limited to the region of convergence". For the experimental evaluation proposed, interval weights are initially defined as wide as necessary (with amplitudes up to 10^6). In addition, the IA solution adopted by these researchers extends to defining an interval version of the gradient descent procedure. On the contrary, the method presented in this paper uses IA concepts only for computing effective intervals for the initial weights and therefore it is not computationally expensive.

The sections of this paper are organized as follows. Section 2 is devoted to a presentation of the IA concepts underpinning the LIT-Approach. Section 3 presents the analysis of LIT-Approach including both theoretical results and the weight initialization algorithm. Section 4 is dedicated to the experimental evaluation of our approach and its comparison with well known initialization procedures. Finally, Section 5 summarizes the paper with some concluding remarks.

2. Interval analysis and the tolerance problem

2.1. Interval arithmetic

The arithmetic defined on sets of intervals, rather than sets of real numbers is called interval arithmetic. An interval or interval number I is a closed interval $[a, b] \subset \mathbb{R}$ of all real numbers between (and including) the endpoints a and b, with $a \leqslant b$. The terms interval number and interval are used interchangeably. Whenever $a = b$ the interval is said to be degenerate, thin or even point interval. An interval X may be also denoted as $\left[\underline{X}, \overline{X}\right]$, $[X]$ or even $[X_L, X_U]$ where subscripts $_L$ and $_U$ stand for lower and upper bounds respectively. Interval variables may be uppercase or lowercase, (Alefeld & Mayer, 2000). In this paper, identifiers for intervals and interval objects (variables or vectors) will be denoted with boldface lowercase such as $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ and boldface uppercase notation will be used for matrices, e.g. $\boldsymbol{X}$. Lowercase letters will be used for the square bracketed notation of intervals $[\underline{x}, \overline{x}]$, or the elements of an interval as a set. An interval $[\underline{x}, \overline{x}]$ where $\underline{x} = -\overline{x}$ is called a symmetric interval. Finally, if $\boldsymbol{x} = [\underline{x}, \overline{x}]$ then the following notation will be used in this paper.

$\mathrm{rad}(\boldsymbol{x}) = (\overline{x} - \underline{x})/2$, is the radius of the interval $\boldsymbol{x}$

$\mathrm{mid}(\boldsymbol{x}) = (\overline{x} + \underline{x})/2$, is the midpoint (meanvalue) of the interval $\boldsymbol{x}$

$|\boldsymbol{x}| = \max\{|\overline{x}|, |\underline{x}|\}$, is the absolute value (magnitude) of the interval $\boldsymbol{x}$

$\mathbb{IR}$, denotes the set of real intervals

$\mathbb{IR}^n$, denotes the set of n-dimensional vectors of real intervals

Let $\diamond$ denote one of the elementary arithmetic operators $\{+, -, \times, \div\}$ for the simple arithmetic of real numbers x, y. If $\boldsymbol{x}, \boldsymbol{y}$ denote real intervals then the four elementary arithmetic operations are defined by the rule

$$\boldsymbol{x} \diamond \boldsymbol{y} = \{x \diamond y \mid x \in \boldsymbol{x},\ y \in \boldsymbol{y}\}. \tag{1}$$

This definition guarantees that $x \diamond y \in \boldsymbol{x} \diamond \boldsymbol{y}$ for any arithmetic operator and any values of x and y. In practical calculations each interval arithmetic operation is reduced to operations between real numbers. If $\boldsymbol{x} = [\underline{x}, \overline{x}]$ and $\boldsymbol{y} = [\underline{y}, \overline{y}]$ then it can be shown that the above definition produces the following intervals for each arithmetic operation:

$$\boldsymbol{x} + \boldsymbol{y} = [\underline{x} + \underline{y}, \overline{x} + \overline{y}] \tag{2a}$$

$$\boldsymbol{x} - \boldsymbol{y} = [\underline{x} - \overline{y}, \overline{x} - \underline{y}] \tag{2b}$$