Pattern Detection with Growing Neural Networks – An Application to Marketing and Library Data

Reinhold Decker and Antonia Hermelbracht

Department of Economics and Business Administration
University of Bielefeld, 33615 Bielefeld, Germany

Abstract. This paper introduces a new growing neural network for pattern detection which bears certain resemblances to the growing neural gas network suggested by Fritzke (1995) [2]. However, the algorithm at hand is more parsimonious with respect to the number of parameters to be specified a priori. Thus it is largely autonomous regarding the data-driven construction of the final network topology which unburdens the user significantly. To demonstrate its performance and adaptability the new algorithm is applied to real classification tasks in lifestyle analysis and media usage analysis.

1 Introduction

In the last two decades self-organizing neural networks have become an important part of the data analytical instruments in natural and social sciences. Recent improvements aim both at the elimination of existing methodological problems and the alignment of available algorithms to specific areas of application. Even though diverse methodological difficulties have been solved by algorithms such as the neural gas network (Martinetz, Schulten 1991) [1], the growing neural gas network (Fritzke 1995) [2], the growing hierarchical self-organizing map (Dittenbach et al. 2002) [3], as well as the grow when required neural network (Marsland et al. 2002) [4], the efficient determination of adequate parameter settings still continues to be a crucial problem.

Against this background the paper at hand presents a contribution to the solution of this problem. Specifically, we are going to introduce a self-organizing neural network where the number of parameters to be preset by the user at the beginning of the adaptation process is limited to two. The so-called Growing Neural Network with Autonomous Parameter Selection (GNNAPS) determines all controlling parameters apart from the data compression level and the number of iterations required autonomously.

The practical benefits of this approach are demonstrated by means of two real data sets covering different areas of application. Firstly, we are going to detect consumer types in the framework of lifestyle analysis. The reliable identification of consumption patterns requires special attention in marketing due to the interrelations between purchase intention and purchase behaviour. Secondly, the detection of media usage patterns in academic libraries will be

shown. Shrinking budgets and increasing costs cause the library managements more and more to adopt marketing research methods to attain deeper understanding of book usage behaviour. The data-driven determination of meaningful book usage patterns may be helpful with regard to the optimization of future media procurement and the adaptation of current services to changing user needs in academic as well as public libraries.

The remainder of this paper is structured as follows. In the next section we give a brief description of the main components of the new algorithm, followed by sketchings of selected findings of both empirical applications. The paper concludes with a brief review of the achieved cognitions.

2 A growing neural network with autonomous parameter selection

The methodology of data compression and feature extraction underlying the new algorithm is vector quantization, the basic idea of which is to represent J input vectors $\boldsymbol{t}_j = (t_{j1}, \dots, t_{jk}, \dots, t_{jK})$ (with $j \in \{1, \dots, J\}$) of a data set Ω by an as small as possible number of weight vectors $\boldsymbol{\eta}_h = (\eta_{h1}, \dots, \eta_{hk}, \dots, \eta_{hK})$ (with $h \in \{1, \dots, H\}$ and $H < J$). To this end, Ω is divided into H classes Ω_h, where each class is represented by a weight vector $\boldsymbol{\eta}_h$. Each input vector $\boldsymbol{t}_j$ is classified into one class Ω_h, or rather assigned to one weight vector $\boldsymbol{\eta}_h$, such that the distance between the input vector and the respective weight vector is minimal (Kohonen 2001) [5]. The only decisions the user has to make in advance when applying the algorithm, apart from indicating the dimensionality K of the data, concern the required data compression level CL $(0 < CL < 1)$ and the maximum number of iterations L (with index $l \in \{1, \dots, L\}$) considered to be necessary.

At the beginning of the adaptation process, i.e. in iteration $l = 1$, the neural network, or rather the associated set of units $\mathcal{U}$, contains two non-connected units u_1 and u_2, i.e. $\mathcal{U} = \{u_1, u_2\}$, with associated weight vectors $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, both initialized with positive random values. The firing counters and the training requirements of these units are initialized according to $y_1 = y_2 = 0$ and $w_1 = w_2 = 1$. The set $\mathcal{C}$ of connections between units is empty. Both sets together determine the topological structure of the neural network.

Furthermore, let S_{Max} be the maximum Euclidean distance between two input vectors and S_{Min} the corresponding minimum. The decision as to whether or not a new unit should be added to the network during an iteration is assumed to depend on the activity of the best matching ('winning') unit, i.e. the one with the smallest distance from the current input vector (Marsland et al. 2002) [4]. The smaller this distance, the higher the activity is. If the activity of the winning unit falls below the threshold $v_{Thres} = \exp\big(-\big(\frac{S_{Max}}{2} - \big(\frac{S_{Max}}{2} - S_{Min}\big) \cdot (1 - CL)^{\frac{1}{4}}\big)\big)$, we take this as a hint at an insufficient fit between the winning unit and the current input vector. Finally, we define learning rates $\epsilon_{Best} = 0.1 + \frac{1}{2}\big(1 - \exp\big(-\frac{CL \cdot S_{Max}}{\sqrt{L}}\big)\big)$, with