Pattern Recognition 47 (2014) 96–113

Contents lists available at ScienceDirect

Pattern Recognition

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

Self-organizing maps whose topologies can be learned with adaptive binary search trees using conditional rotations

César A. Astudillo [a,*], B. John Oommen [b,1]

[a] Department of Computer Science, Universidad de Talca, Merced 437 Curicó, Chile
[b] School of Computer Science, Carleton University, Ottawa, Canada K1S 5B6

ARTICLE INFO

Available online 25 April 2013

Keywords:
Adaptive data structures
Binary search trees
Self-organizing maps

ABSTRACT

Numerous variants of Self-Organizing Maps (SOMs) have been proposed in the literature, including those which also possess an underlying structure, and in some cases, this structure itself can be defined by the user. Although the concepts of growing the SOM and updating it have been studied, the whole issue of using a self-organizing Adaptive Data Structure (ADS) to further enhance the properties of the underlying SOM, has been unexplored. In an earlier work, we impose an *arbitrary*, user-defined, tree-like topology onto the codebooks, which consequently enforced a neighborhood phenomenon and the so-called *tree-based* Bubble of Activity (BoA). In this paper, we consider how the underlying tree itself can be rendered *dynamic* and *adaptively transformed*. To do this, we present methods by which a SOM with an underlying Binary Search Tree (BST) structure can be adaptively re-structured using Conditional Rotations (CONROT). These rotations on the nodes of the tree are local, can be done in constant time, and performed so as to decrease the Weighted Path Length (WPL) of the entire tree. In doing this, we introduce the pioneering concept referred to as *Neural Promotion*, where neurons gain prominence in the Neural Network (NN) as their significance increases. We are not aware of any research which deals with the issue of Neural Promotion. The advantage of such a scheme is that the user need not be aware of any of the topological peculiarities of the stochastic data distribution. Rather, the algorithm, referred to as the TTOSOM with Conditional Rotations (TTOCONROT), converges in such a manner that the neurons are ultimately placed in the input space so as to represent its stochastic distribution, and additionally, the neighborhood properties of the neurons suit the best BST that represents the data. These properties have been confirmed by our experimental results on a variety of data sets. We submit that all these concepts are both novel and of a pioneering sort.

1. Introduction

This paper is a pioneering attempt to merge the areas of Self-Organizing Maps (SOMs) with the theory of Adaptive Data Structures (ADSs). Put in a nutshell, we can describe the goal of this paper as follows: Consider a SOM with n neurons. Rather than having the neurons merely possess information about the feature space, we also attempt to *link* them together by means of an underlying Data Structure (DS). This DS could be a singly-linked list, a doubly-linked list or a Binary Search Tree (BST), etc. The intention is that the neurons are governed by the laws of the SOM *and* the underlying DS. Observe now that the concepts of "neighborhood" and Bubble of Activity (BoA) are not based on the nearness of the neurons in the feature space, but rather on their proximity in the underlying DS. Having accepted the above-mentioned premise, we intent to take this entire concept to a higher level of abstraction and propose to modify this DS *itself* adaptively using operations specific to it. As far as we know, the combination of these concepts has been unreported in the literature.

Before we proceed, to place our results in the right perspective, it is probably wise to see how the concept of neighborhood has been defined in the SOM literature.

Kohonen, in his book [36], mentions that it is possible to distinguish between two basic types of neighborhood functions. The first family involves a kernel function (which is usually of a Gaussian nature). The second, is the so-called neighborhood set, also known as the Bubble of Activity (BoA). This paper focuses on the second type of neighborhood function.

Even though the traditional SOM is dependent on the neural distance to estimate the subset of neurons to be incorporated into

* Corresponding author. Tel.: +56 9 76971862.
E-mail addresses: castudillo@utalca.cl (C.A. Astudillo), oommen@scs.carleton.ca (B.J. Oommen).

[1] This author is also an *Adjunct Professor* with the University of Agder in Grimstad, Norway. The work of this author was partially supported by NSERC, the Natural Sciences and Engineering Research Council of Canada.

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.04.012

the BoA, this is not always the case for the SOM-variants included in the literature. Indeed, the different strategies described in the state-of-the-art utilize families of schemes to define the BoA. We mainly identify three sub-classes.

The first type of BoA uses the concept of the neural distance as in the case of the traditional SOM. Once the Best Matching Unit (BMU) is identified, the neural distance is calculated by traversing the underlying structure that holds the neurons. An important property of the neural distance between two neurons is that it is proportional to the number of connections separating them. Examples of strategies that use the neural distance to determine the BoA are the Growing Cell Structures (GCS) [24], the Growing Grid (GG) [25], the Incremental Grid Growing (IGG) [13], the Growing SOM (GSOM) [3], the Tree-Structured SOM (TSSOM) [37], the Hierarchical Feature Map (HFM) [43], the Growing Hierarchical SOM (GHSOM) [50], the Self-Organizing Tree Algorithm (SOTA) [22], the Evolving Tree (ET) [46], the Tree-based Topology Oriented SOM (TTOSOM) [8], among others.

A second subset of strategies employ a scheme for determining the BoA that does not depend on the inter-neural connections. Instead, such strategies utilize the distance in the feature space. In these cases, it is possible to distinguish between two types of Neural Networks (NNs). The simplest situation occurs when the BoA only considers the BMU, i.e., it constitutes an instance of hard Competitive Learning (CL), as in the case of the Tree-Structured VA (TSVQ) [37] and the Self-Organizing Tree Map (SOTM) [27].

A more sophisticated and computationally expensive scheme involves ranking the neurons as per their respective distances to the stimulus. In this scenario, the BoA is determined by selecting a subset of the closest neurons. An example of a SOM variant that uses such a ranking is the Neural Gas (NG) [40].

According to the authors of [46], the SOM-based variants included in the literature attempt to tackle two main goals: They either try to design a more flexible topology, which is usually useful to analyze large data sets, or to reduce the most time-consuming task required by the SOM, namely, the search for the BMU when the input set has a complex nature. In this paper we focus on the former of the two mentioned goals. In other words, our goal is to enhance the capabilities of the original SOM algorithm so as to represent the underlying data distribution and its structure in a more accurate manner. We also intend to do so by constraining the neurons so that they are related to each other, *not just based on their neural indices and stochastic distribution*, but also based on a BST relationship.

Furthermore, as a long term ambition, we also anticipate methods which can accelerate the task of locating the nearest neuron during the CL phase. This work will present the details of the design and implementation of how an adaptive process applied to the BST, can be integrated into the SOM.

Regardless of the fact that numerous variants of the SOM have been devised, few of them possess the ability of modifying the underlying topology [13,21,22,26,27,42,46,52]. Moreover, only a small subset use a tree as their underlying DS [8,21,22,27,46,52]. These strategies attempt to dynamically modify the nodes of the SOM, for example, by adding nodes, which can be a single neuron or a layer of a SOM-grid. However, our hypothesis is that it is also possible to attain to a better understanding of the unknown data distribution by performing *structural* tree-based modifications on the tree, which although they preserve the general topology, attempt to modify the overall configuration, i.e., by altering the way by which nodes are *interconnected*, and yet continue as a BST. We accomplish this by dynamically adapting the edges that connect the neurons, by rotating[2] the nodes within the BST that holds the whole structure of neurons. As we will explain later, this is further achieved by local modifications to the overall structure in a constant number of steps. Thus, we attempt to use rotations, tree-based neighbors *and* the feature space to improve the quality of the SOM.

1.1. Motivations

Acquiring information about a set of stimuli in an unsupervised manner, usually demands the deduction of its structure. In general, the *topology* employed by any Artificial Neural Network (ANN) possessing this ability has an important impact on the manner by which it will “absorb” and display the properties of the input set. Consider for example, the following: A user may want to devise an algorithm that is capable of learning a triangle-shaped distribution as the one depicted in Fig. 1. The SOM tries to achieve this by defining an underlying grid-based topology and to fit the grid within the overall shape, as shown in Fig. 1a (duplicated from [36]). However, from our perspective, a grid-like topology does not naturally fit a triangular-shaped distribution, and thus, one experiences a deformation of the original lattice during the modeling phase. As opposed to this, Fig. 1b shows the result of applying one of the techniques developed by us, namely the TTOSOM [8]. As the reader can observe from Fig. 1b, a 3-ary tree seems to be a far more superior choice for representing the particular shape in question.

On closer inspection, Fig. 1b depicts how the complete tree fills in the triangle formed by the set of stimuli, and further, seems to do it *uniformly*. The final position of the nodes of the tree suggests that the underlying structure of the data distribution corresponds to the triangle. Additionally, the root of the tree is placed roughly in the center of mass of the triangle. It is also interesting to note that each of the three main branches of the tree, cover the areas directed towards a vertex of the triangle respectively, and their sub-branches fill in the surrounding space around them in a recursive manner, which we identify as being a holograph-like behavior.

Of course, the triangle of Fig. 1b serves only as a very simple *prima facie* example to demonstrate to the reader, in an informal manner, how both techniques will try to learn the set of stimuli. Indeed, in real-world problems, these techniques can be employed to extract the properties of high-dimensional samples.

One can argue that imposing an initial topological configuration is not in accordance with the founding principles of unsupervised learning, the phenomenon that is supposed to occur without “supervision” within the human brain. As an initial response we argue that this “supervision” is required to enhance the *training* phase, while the information we provide relates to the *initialization* phase. Indeed, this is in line with the well-accepted principle [23], that very little can be automatically learned about a data distribution if no assumptions are made!

As the next step of motivating this research endeavor, we venture into a world where the neural topology *and* structure are themselves learned during the training process. This is achieved by the method that we propose in this paper, namely the TTOSOM with Conditional Rotations (TTOCONROT), which, in essence, dynamically extends the properties of the above-mentioned TTOSOM. Again, to accomplish this we need key concepts that are completely new to the field of SOMs, namely those related to tree-based Adaptive Data Structure (ADS). Indeed, as demonstrated by our experiments, the results that we have already obtained have been applauded by the research community,[3] and these, to the best of our knowledge, have remained unreported in the literature.

[2] The operation of rotation is the one associated with BSTs, as will be presently explained.

[3] As mentioned earlier, a paper which reported the preliminary results of this study, won the *Best Paper Award* in a well-known international AI conference [7].