AC-ViSOM: Hybridising the Modified Adaptive Coordinate (AC) and ViSOM for Data Visualization

Md. Sarwar Zahan Tapan
Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak (UNIMAS)
94300, Kota Samarahan, Sarawak, Malaysia
E-mail: sarwar740@yahoo.com

Teh Chee Siong
Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak (UNIMAS)
94300, Kota Samarahan, Sarawak, Malaysia
E-mail: csteh@fcs.unimas.my

Abstract

ViSOM's (Visualization induced SOM) final map can be seen as a smooth net embedded in the input space, where the distances among neurons are controlled by a regularization control parameter which is usually heuristically chosen on a trial and error basis. Empirical studies shown that ViSOM suffers from dead neuron problem, since a big number of neurons fall outside of the data region due to the regularization effect, even though the regularization control parameter is properly chosen. In this paper, a modified Adaptive Coordinate (AC) approach that is able to preserve data structure is hybridised with ViSOM is proposed. Experimental studies on benchmark datasets shown that the proposed method was able to eliminate the selection of regularization control parameter and minimizing the dead neuron for better data representation.

1. Introduction

Self-Organizing Map (SOM) [1] is able to preserve data topology in its visualization from N-dimensional space to the low dimensional display space. However, SOM's visualization is not directly able to represent the data structure and inter-neuron distance in its topology preserved mapping [2]. Visualization induced SOM (ViSOM) [3] is then proposed to enhance SOM's visualization to preserve data structure and inter-neuron distances. Empirical studies in [3] have proven ViSOM's superiority among several other visualization methods such as Principal Component Analysis (PCA) [4], Multidimensional Scaling (MDS) [5] and Sammon's Mapping [6] in terms of data structure, data topology and inter-neuron distance preservation. proposed in [9][10] to produce data structure and inter-neuron preserved topology mapping.

ViSOM's final map can be seen as a smooth net (mesh grid) embedded in the input space, where the distances among neurons are controlled by a regularization control parameter, which is usually heuristically chosen using trial and error approach [2]. This regularization control parameter controls the regular spread of the weight vectors in the data region in N-dimensional space. Empirical studies in [3] have shown that ViSOM suffers from dead neuron problem, since a big number of neurons fall outside of the data region in the produced map due to the regularization impact even though the regularization control parameter value is reasonably selected [3].

Empirical studies in [3] have also shown that, for a suitable regularization control parameter, ViSOM requires a large amount of neurons in the network to effectively represent the data. The big number of dead neurons produced in the ViSOM map results non-optimised network resource utilization even though the larger map size caused more computation efforts. Thus, this work scopes itself in defining a way of eliminating the impact of the regularization control parameter in the ViSOM network and optimise the dead-neuron problem, at the same time produce data structure, data topology and inter-neuron distance preserved data visualization.

In this work, a modified Adaptive Coordinate (AC) approach is hybridised with ViSOM to address the above-mentioned issues. The basic idea of AC approach [7][8] is to mirror the movements and locations of the neurons' weight vectors in the high dimensional input space in a low dimensional output space to reveal the clustering tendency of data learned by SOM [7]. Usually the AC is applied on the topology preserved SOM grid to reveal the clustering tendency of data by removing the rigidity of SOM grid [8]. A modified AC approach is also

In this paper, the integration of the modified AC and ViSOM network is proposed as AC-ViSOM. Section 2 describes ViSOM in brief. Section 3