A Hybrid Supervised ANN for Classification and Data Visualization

Chee Siong Teh and Md. Sarwar Zahan Tapan

Abstract— **Supervised ANNs such as Learning Vector Quantization (LVQs) and Multi-Layer Perceptrons (MLPs) usually do not support data visualization beside classification. Unsupervised visualization focused ANNs such as Self-organizing Maps (SOM) and its variants such as Visualization induced SOM (ViSOM) on the other hand, usually do not optimize data classification as compared with supervised ANNs such as LVQ. Thus to provide supervised classification and data visualization simultaneously, this work is motivated to propose a novel hybrid supervised ANN of LVQwithAC by hybridizing LVQ and modified Adaptive Coordinate (AC) approach. Empirical studies on benchmark data sets proven that, LVQwithAC was able to provide superior classification accuracy than SOM and ViSOM. Beside LVQwithAC was able to provide data topology, data structure, and inter-neuron distance preserve visualization. LVQwithAC was also proven able to perform promising classification among other supervised classifiers besides its additional data visualization ability over them. Thus, for applications requiring data visualization and classification LVQwithAC demonstrated its potential if supervised learning is all possible.**

I. Introduction

ANNs usually do not support supervised classification and data visualization simultaneously. Supervised ANNs, such as Learning Vector Quantization algorithms (LVQs) and Multi-Layer Perceptrons (MLPs) perform classification without providing visualizations [1].

Data visualization on the other hand, has been supported by various approaches such as Principal Component Analysis (PCA) [3], Multidimensional Scaling (MDS) [4], Sammon's Mapping [5], Principal Curves [6] and Principal Surfaces [7]. PCA, MDS and Sammon's Mapping are not proven efficient for data visualization purposes since they demonstrate some major disadvantages such as PCA loses certain useful information during the dimension reduction while MDS and Sammon's require very heavy computation and eventually become impractical for practical applications. Moreover, MDS and Sammon's Mapping cannot accommodate new data sample in the visualization without re-computing the existing data samples [8].

ANN methods such as Self-Organizing Maps (SOM) [2] and its recent variant such as Visualization induced SOM (ViSOM) [9], Probabilistic Regularized SOM (PRSOM) [8] are able to overcome the above mentioned limitations of PCA, MDS and Sammon's Mapping.

Manuscript received Nov 29, 2007. This work was supported by Universiti Malaysia Sarawak (UNIMAS), Malaysia.

C. S. Teh and Z. T. Sarwar are with the Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300, Malaysia. (phone:+6082 581572; fax: +6082 581567; e-mail: csteh@fcs.unimas.my, sarwar740@yahoo.com).

ViSOM [9] and PRSOM [8] have been proposed to enhance SOM's visualization to preserve the inter-neuron distances from *N*-dimensional input space to low dimensional output space. In [8] and [9], PRSOM and ViSOM have already been proven able to provide better data visualization than that of PCA, MDS, Sammon's Mapping and SOM in terms of inter-neuron distance, data topology and data structure preservation.

These visualization focused unsupervised ANNs however do not optimize classification performance compared against supervised classification methods such as LVQ. Thus, statistical classification problems are recommended to be addressed with supervised methods for better classification accuracies if supervised learning is all possible [2]. The supervised classification methods however usually do not provide data visualization at the same time. Thus, this work focuses on developing a hybrid supervised ANN method to perform supervised classification and data visualization simultaneously.

In this work, the Adaptive Coordinate (AC) [10][11] approach is investigated to be hybridized with LVQ to address the above-mentioned issues. The AC approach is an extension of the learning procedure of SOM. The basic idea of AC [10] approach 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.

In this paper, integration between LVQ and a modified Adaptive Coordinates approach is proposed as LVQwithAC. In this proposed method, LVQ performs classification task while a modified AC component projects the data visualization in a predefined 2D output space.

Section 2 of this paper describes LVQ and Adaptive Coordinates (AC) with proposed modification in adaptation criteria. In section 3, integration between LVQ and AC (LVQwithAC) is presented. In section 4, experimental results demonstrate that, LVQwithAC is able to provide simultaneous data visualization and classification. Conclusions are drawn in section 5.

II. LVQ, AC and Proposed Modification in AC

A. Learning Vector Quantization (LVQ)

LVQ is a supervised vector quantization algorithm proposed by Kohonen [2] for statistical classification. It signifies a class of related algorithms such as LVQ1, LVQ2, LVQ2.1, LVQ3 and OLVQ1 [2]. Let assume that, all samples of $\boldsymbol{x}$ are derived from a finite set of classes $\{S_k\}$ with overlapping distributions. Initially several codebook vectors are assigned to each class of $\boldsymbol{x}$ values. Then for each data sample $\boldsymbol{x}(t)$ in discrete time step t, the winner neuron $\boldsymbol{m}_c(t)$ is

978-1-4244-1821-3/08/$25.00©2008 IEEE