Hybridization of Learning Vector Quantization (LVQ) and Adaptive Coordinates (AC) for data classification and visualization

Md. Sarwar Zahan Tapan and Chee Siong Teh
Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak (UNIMAS)
94300, Kota Samarahan, Sarawak, Malaysia
e-mail: sarwar740@yahoo.com, csteh@fcs.unimas.my

***Abstract*—Most of the Artificial Neural Network (ANN) methods do not support data classification and visualization simultaneously. Some ANN methods such as Learning Vector Quantization (LVQ), Multi-Layer Perceptrons (MLP) and Radial Basis Function (RBF) perform classification without any visualization. Excellent data visualization on the other hand has been prominently supported by various unsupervised methods such as Self-Organizing Maps (SOM) and its recent variants of Visualization induced SOM (ViSOM) and Probabilistic Regularized SOM (PRSOM). However, being unsupervised these methods do not optimize classification accuracy compared with the supervised classification methods such as LVQ. Thus, the scope of a novel supervised method is felt necessary to facilitate applications requiring good data visualization and intensive classification. LVQ demonstrates classification performance at least as high as other supervised ANN classifiers. Adaptive Coordinate (AC) on the other hand, has demonstrated the ability of mirroring weight vectors' movements in *N*-dimensional input space to low dimensional output space to reveal the clustering tendency of data learned by SOM. This mirroring concept motivates this work to hybridize a modified AC with LVQ (LVQwihAC) to support data visualization and classification simultaneously. Empirical studies on benchmark data sets demonstrated that, the LVQwihAC method provides better classification accuracy than the unsupervised methods of SOM, ViSOM and PRSOM besides its promising data visualization with higher computational efficiency. The classification performance is also found at least as good as other supervised classifiers with additional data visualization abilities over them.**

I. INTRODUCTION

Most of the Artificial Neural Network (ANN) methods do not support data classification and visualization simultaneously. Some ANN methods such as Learning Vector Quantization (LVQ), Multi-Layer Perceptrons (MLP) and Radial Basis Function (RBF) perform classification without providing visualizations [1]. Learning Vector Quantization (LVQ) is a neural network based classification focused supervised competitive learning model [2]. The codebook vectors of LVQ are placed in the input space and updated according to the vector-learning rules with no data visualization abilities. The classification performance of LVQ is at least as high as other ANN based classification models [2].

On the other hand, data topology preservation and visualization has been supported by various approaches such as Principal Component Analysis (PCA) [3], Multidimensional Scaling (MDS) [4], Sammon's Mapping [5], Principal Curves [6], Principal Surfaces [7], Curvilinear Component Analysis (CCA) [8] and Self-Organizing Maps (SOM) [2]. These methods have been used for dimensionality reduction, data clustering tendency analysis, data topology preservation and projection [8]. Visualization induced SOM (ViSOM) [9] and Probabilistic Regularized SOM (PRSOM) [8] have been proposed to enhance SOM's visualization to preserve the data structure and inter-neuron distances. In [8] and [9], PRSOM and ViSOM have already been proven superior among the other visualization methods. However, being unsupervised, these visualization-focused methods do not optimize classification performance compared with the supervised classification methods such as LVQ. As Kohonen stated in [2], supervised methods are facilitated with higher classification abilities over the unsupervised methods and eventually statistical classification problems are recommended to be addressed with supervised methods for better classification accuracies if supervised learning is all possible [2].

However, the unsupervised data visualization methods do not optimize classification accuracy compared with supervised classification methods, while the supervised classification methods usually do not provide data visualization. Thus the scope of a novel supervised method is felt necessary to facilitate applications requiring good data visualization and intensive classification. This research thus focuses on developing a hybrid supervised ANN method to perform data classification and data visualization simultaneously.

The prominent data visualization methods of PRSOM and ViSOM's final map can be seen as a smooth net (mesh grid) embedded in the input space, where the distances between each two neighboring neurons is controlled by a regularization control parameter [8]. For a suitable regularization control parameter value, both of these methods require a large amount of neurons in the network to represent the data, hence required computation is higher [8]. Empirical studies in [8] and [9] have shown that both of these methods are vulnerable to suffer

1-4244-1355-9/07/$25.00 @2007 IEEE