

ABSTRACT

The Self Organizing Map (SOM) proposed by T.Kohonen (1982), has been widely used in industrial applications such as pattern recognition, biological modelling, data compression, signal processing and data mining (T. Kohonen, 1997; M.N.M Sap and E. Mohebi, 2008a, 2008b, 2008c). It is an unsupervised and nonparametric neural network approach. The success of the SOM algorithm lies in its simplicity that makes it easy to understand, simulate and be used in many applications. The basic SOM consists of neurons usually arranged in a two-dimensional structure such that there are neighbourhood relations among the neurons. After completion of training, each neuron is attached to a feature vector of the same dimension as input space. By assigning each input vector to the neuron with nearest feature vectors, the SOM is able to divide the input space into regions (clusters) with common nearest feature vectors. This process can be considered as performing vector quantization (VQ) (R.M. Gray, 1984). In addition, because of the neighborhood relation contributed by the inter-connections among neurons, the SOM exhibits another important property of topology preservation. Clustering algorithms attempt to organize unlabeled input vectors into clusters such that points within the cluster are more similar to each other than vectors belonging to different clusters (N. R. Pal, et al., 1993). The clustering methods are of five types: hierarchical clustering, partitioning clustering, density-based clustering, grid-based clustering and model-based clustering (J. Han and M. Kamber, 2000). The rough set theory employs two upper and lower thresholds in the clustering process, which result in a rough clusters appearance. This technique also could be defined in incremental order i.e. the number of clusters is not predefined by users. In this chapter, a new two-level clustering algorithm is proposed. The idea is that the first level is to train the data by the SOM neural network and then clustering at the second level is a rough set based incremental clustering approach (S. Ashraf, et al., 2006), which will be applied on the output of SOM and requires only a single neurons scan. The optimal number of clusters can be found by rough set theory, which groups the given neurons into a set of overlapping clusters (clusters the mapped data respectively). Then the overlapped neurons will be assigned to the true clusters they belong to, by apply simulated annealing algorithm. A simulated annealing algorithm has been adopted to minimize the uncertainty that comes from some clustering operations. In our previous work (M.N.M. Sap and E. Mohebi, 2008a) the hybrid SOM and rough set has been applied to catch the overlapped data only, but the experiment results show that the proposed algorithm (SA-Rough SOM) outperforms the previous one. This chapter is organized as following; in section 2, the basics of SOM algorithm are outlined. The Incremental Clustering and Rough set theory are described in section 3. In section 4, the essence of simulated annealing is described. The proposed algorithm is presented in section 5. Section 6 is dedicated to experiment results, section 7 provides brief conclusion, and future works and an outline of the chapter summary is described in section 8.