402a Self- Constructing and Organizing Neural Network Based Fuzzy Clustering Technique and Its Application to Fault Diagnosis

Bharat Bhushan and Jose A. Romagnoli

The advent of faster and more reliable computer systems has revolutionized the manner in which industrial processes are monitored and controlled. These advances have resulted in the generation of a large amount of process data, yet the task of interpreting and analysing this data is daunting. Fault diagnosis is one of the most important aspects of process monitoring. It is closely related to the pattern recognition task. The entire task of the pattern recognition can be divided into three phases: data acquisition and pre-processing, feature extraction and pattern classification.

The data generated in a process are high dimensional and strongly correlated. For efficient analysis of the process condition, these data must be projected to a lower dimensional space with the least loss of the information. Dimensionality reduction can be closely related to feature extraction in which the feature space contains the derived variables, function of original process variables, which efficiently capture the information in the original data. Furthermore, each data points of the feature space are projected to the classification space. The pattern belonging to the same class will cluster together. Each cluster represents a distinct class, and cluster of points represent different classes of patterns.

Several methods have been suggested for pattern classification and clustering. In most of the classification techniques, the size of the network is predetermined. On the other hand, in most of the cases, a priori knowledge to decide the suitable size is not available. The appropriate size can only be achieved through many simulations. Most of the clustering techniques need a priori information about the number of clusters. The performance of the clustering techniques based on objective function strongly depends on the given initial solution and in many cases; the solution is trapped in the local optima. Genetic algorithm is one of the most efficient techniques for global optimization; however, it is computationally expensive. In addition, the boundaries between the two classes are sharp i.e. either a point belongs to a class or not. However, in real applications there is very often no sharp boundary between the classes.

In this work, a novel methodology for pattern classification based on unsupervised learning is proposed. The size of the network is determined incrementally during the simulation. Each node of the network is described by a multidimensional Gaussian function. When a training data is presented to the network, the membership value of data point in each of the node is found. The centre of the nodes and deviation around the node is learnt and the nodes relocate it accordingly. Each data points in the feature space have a membership degree between zero and one to the nodes. The membership degree is used as the information to decide whether to generate a new node or not. These steps are repeated until certain termination criterions are met. It is observed that similar feature vector has high membership degree with the close nodes and the closer nodes have similar feature vectors. The distribution of the nodes in the classification space is closely related to the probability density of the feature space. The region where the probability distribution is high has higher number of nodes compared to the lower density region.

This methodology is further enhanced for fuzzy clustering. The advantage of the proposed method over the existing clustering techniques is its ability to automatically determine the number of clusters. Since the methodology is based on learning, it is computationally less expensive and the result is not affected by the initial guesses. The methodology is quite generic in nature and can be used for pattern classification, clustering, and data visualization.

To test the overall strategy, a general-purpose pilot facility is used. This pilot plant is located in Process System Engineering Laboratory at the University of Sydney, Australia. This unit contains two CSTRs, a mixer, a feed tank and a number of heat exchangers. Each CSTR consists of a reaction vessel, a steam jacket, a cooling coil, and a stirrer. Material from the feed tank is heated before being fed to the first reactor and the mixer. The effluent from the first reactor is then mixed with the material in the material in the mixer before being fed to the second reactor. The effluent from the second reactor is, fed back to the feed tank and the cycle continues. The pilot plant is well instrumented to provide many possible control scenarios and configurations.

Different type of faulty conditions was simulated. Principal component analysis is one of the most used and effective method for feature extraction. The data was projected to a lower dimensional feature space using principal component analysis. Each data point of the feature space is further projected to the classification space using above methodology. There is not a crisp boundary between the classes. Instead, each feature vector has a membership degree in each of the class. The severity of a particular fault is determined by its position from the cluster centres. The result shows that the proposed methodology is quit effective for the fault diagnosis of a chemical process.