ABSTRACT

Classification is the process of assigning a data instance to a class or category based on the characteristics of that data in such a way that all the instances in one class exhibit similar characteristics while data in different classes are unique. To extract information from a large amount of data the first and foremost task is to group the data into classes. Any algorithm or model that performs the task of classification is called a classifier. Classifiers can be designed manually based on expert’s knowledge but with the available technology it is possible to build classifiers that can learn from real data. Building the classifier is a multi-step approach. The classifier is evaluated based on the accuracy, speed, robustness, interpretability and scalability.

In this research work, the classification method used to design the classifier is neural networks. The automobile sales data obtained from Chevrolet Sales and Service, Tirunelveli, is used to train and test the performance of the classifier. Neural Networks is a powerful tool used to build intelligent systems. Radial Basis Function (RBF) neural network with Spherical Gaussian Function (SGF) as the activation function is found to produce excellent results in classifying the given sales data into three classes: high sales, moderate sales and low sales. Compared to other existing classifiers like multi-layer perceptrons, RBF neural classifier produces better accuracy because it performs classification by measuring the similarity of the input to a prototype selected for each class.

The accuracy and speed of the classifier is greatly affected by the noisy and irrelevant data instances. Dimensionality reduction is the technique used to remove noisy and redundant features in the data set. Feature extraction and feature selection are the major components of the dimensionality reduction technique. Feature selection
selects a subset of features relevant to the problem under study whereas feature extraction creates a new set of instances eliminating redundancy. In this research work, correlation techniques and clustering analysis are applied to remove noisy data and to extract relevant instances from the original data set. K-means clustering algorithm is used to form the initial clusters based on the similarities in the input data set which is further grouped into mutually exclusive clusters using hierarchical agglomerative clustering algorithm. This approach is found to be effective in producing a faster neural classifier.

The hybridization of fuzzy logic into neural classification produces a highly intelligent system that incorporates the human-like reasoning style. Fuzzy C-means clustering algorithm is used to select the initial clusters instead of k-means clustering as it suffers from the random selection of the initial prototype. This resulted in a far better classifier with a good classification accuracy of about 95%. To fine tune the performance of the classifier, a new membership function called Radial Cubic B-spline Membership Function (RCBMF) is used define the degree of membership of each input data instance. RCBMF-RBF neural classifier model is trained and tested with the original sales data. The performance of this model is compared with that of the other fuzzy neural classifiers using traditional trapezoidal membership functions. The experimental results show that the RCBMF-RBF neural classifier is found to be efficient and faster in classifying the given data set with a classification accuracy of 99% and the average training time required is 1.27 seconds.