The Use of a Bayesian Neural Network Model for Classification Tasks

Abstract

This thesis deals with a Bayesian neural network model. The focus is on how to use the model for automatic classification, i.e. on how to train the neural network to classify objects from some domain, given a database of labeled examples from the domain. The original Bayesian neural network is a one-layer network implementing a naive Bayesian classifier. It is based on the assumption that different attributes of the objects appear independent of each other. This work has been aimed at extending the original Bayesian neural network model, mainly focusing on three different aspects.

First the model is extended to a multi-layer network, to relax the independence requirement. This is done by introducing a hidden layer of complex columns, groups of units which take input from the same set of input attributes. Two different types of complex column structures in the hidden layer are studied and compared. An information theoretic measure is used to decide which input attributes to consider together in complex columns. Also used are ideas from Bayesian statistics, as a means to estimate the probabilities from data which are required to set up the weights and biases in the neural network.

The use of uncertain evidence and continuous valued attributes in the Bayesian neural network are also treated. Both things require the network to handle graded inputs, i.e. probability distributions over some discrete attributes given as input. Continuous valued attributes can then be handled by using mixture models. In effect, each mixture model converts a set of continuous valued inputs to a discrete number of probabilities for the component densities in the mixture model.

Finally a query-reply system based on the Bayesian neural network is described. It constitutes a kind of expert system shell on top of the network. Rather than requiring all attributes to be given at once, the system can ask for the attributes relevant for the classification. Information theory is used to select the attributes to ask for. The system also offers an explanatory mechanism, which can give simple explanations of the state of the network, in terms of which inputs mean the most for the outputs.

These extensions to the Bayesian neural network model are evaluated on a set of different databases, both realistic and synthetic, and the classification results are compared to those of various other classification methods on the same databases. The conclusion is that the Bayesian neural network model compares favorably to other methods for classification.

In this work much inspiration has been taken from various branches of machine learning. The goal has been to combine the different ideas into one consistent and useful neural network model. A main theme throughout is to utilize independencies between attributes, to decrease the number of free parameters, and thus to increase the generalization capability of the method. Significant contributions are the method used to combine the outputs from mixture models over different subspaces of the domain, and the use of Bayesian estimation of parameters in the expectation maximization method during training of the mixture models.