Predicting CardioVascular Risk Using Neural Net Techniques

T.T.Nguyen and D.N. Davis, Computer Science, University of Hull.

Introduction

Neural Networks are broadly applied in a number of fields such as cognitive science, diagnosis, and forecasting. Medical decision support is one area of increasing research interest. Ongoing collaborations between cardiovascular clinicians and computer science are looking at the application of neural networks (and other data mining techniques) to the area of individual patient diagnosis, based on clinical records (from Hull and Dundee sites). The current research looks to advance initial investigations in a number of ways. Firstly through a deeper analysis of the clinical data, using data mining and statistical tools, we hope to be able to extend the usefulness of much of the clinical data set. Problems with the data include differences in attribute presence and use across different sites, and missing values. Secondly we look to advance the classification of referred patients (into high, medium and low risk) through the rigorous use of both supervised and unsupervised neural net techniques. Through the use of different classifiers, a better clinical diagnostic support model may be built.

Method

These experiments made use of data prepared for an earlier study [1]. The data had been transformed from its original mixed type (categorical, Boolean and continuous) to purely numerical (0 or 1 or continuous in the range 0 to 1). This is appropriate to it being used with neural networks. The original data transformation was subject to a number of minor flaws; these are corrected in this study. We compare the results from the original study with results using the corrected data, with multi-layer perceptrons (MLP), radial basis functions (RBF), support vector machines (SVM) and self organizing feature maps (SOM). Comparisons are made using the Mean Square Error (MSE). 300 data records were used for training purposes; 700 records used for cross-validation. The cross-validation set is used to define the optimal cut-off point for training.

Results

Table 1 shows results obtained using eight different neural techniques. This includes MLP and RBF with zero, one and two hidden layers. These replicate the best performing topologies in the clinical study (given as last row). The SVM and SOM topologies were not used in the initial study. The best performing classifier was the Multi-Layer Perceptron with 1 hidden layer (MLP-1H, MinMSE(CV)= 0.059). Overall there is little variation between the classifiers in terms of their minimum MSE on the validation data, with the exception of the SVM. There is a noticeable difference between the remaining topologies is the training time to convergence.

	MLP-0H	MLP-1H	MLP-2H	RBF-0H	RBF-1H	RBF-2H	SVM	SOM
MinMSE (CV)	0.072	0.059	0.06	0.06	0.06	0.06	0.122	0.06
Epochs	903	96	107	53	899	194	84	52
MSE (T)	0.061	0.048	0.046	0.052	0.052	0.052	0.057	0.052
MSE (study 1)	0.12	0.15	0.13	0.25	0.07	0.09	---	----

Table 1. Comparison of MSE for 8 neural classifiers.

Discussion and Further Work

All the developed classifiers (except SVM) improve on the best results from the initial clinical study, despite in being hampered by reduced training volume. It is suggested that a more rigorous data preparation stage is responsible. How ever, these results need to be subjected to further analysis. One way will be to investigate the confusion matrix for each classifier to see the type of errors being made. A comparison of supervised (MLP/RBF) versus unsupervised (SOM) classifiers may help in determining more appropriate patient classifications. These results can then be applied in determining what of the original data should be used to generate a better set of classifiers of use in predicting CardioVascular risk.

References

[1] Kuhan, G, Davis, DN Chetter, IC, McCollum, CN & McCollum, PT. The use of Artificial Neural Networks for risk prediction following Carotid endarterectomy, Internal Report, University of Hull, 2003.