Background

Data obtained for this case study is obtained from Veteran's Health Administration (VHA) Hospital which is a large tertiary care located metropolitan Chicago area with a 62-acre campus. The VHA is an active participant in the full perpetuity care of Spinal Cord Injury (SCI) patients. This medical facility has a rehabilitation centre and SCI care units with a total of 68 beds, a hospital-based SCI home care program, and a 30-bed residential SCI unit. At a given time there are about 500 to 800 patients without knowing their length of stay.

Brief description

The focus of this article is the prediction of length of stay (LOS) of patients in the facility. Although these patients are not in abundance, they do form outliers in the healthcare system due to their extended hospital stays and comparatively higher treatment costs. Output of this research would help hospitals understand the approx. requirements of the respective patients and allocate enough resources.

Data description

One of the most extensively used medical information system in the world known as the Veterans Health Information Systems and Technology Architecture (VistA) is administered at VHA hospital. The system provides extensive administrative and clinical support along with documentation. The samples used for study were patient admitted from one inpatient SCI Unit. 597 SCI patients with 1107 admissions from this study unit admitted between July 1989 and June 2000 were the chosen study samples. Various noisy data which included missing data such as no follow-ups and mismatched entries were eliminated, leaving behind 525 patients with 1107 admissions.

The 11-year old database had 4750 nursing diagnoses labels that after through evaluation represented 161 unique nursing diagnoses which were further clustered into 20 categories. These 20 categories were used as input variables in the mining software and model predictions.

Algorithms and tools

This case study only used ANNs for building predictive models. Four type of models were developed based on supervised learning with known output (i.e. Length of stay) for comparison.

Dynamic Model

Prune neural networks

Multi-layered perceptron

Radical basis function network

The dynamic model makes an underlying network topology that is changed by including as well as expelling hidden units as the training advances. The Prune neural network is more like decision tree method. It actually, 'prunes' away certain input variables which are dependent on their importance and weight for prediction of the output variable. Model is relatively slower but provides better results. Multilayered Perceptron is one where each hidden layer contains weighted combinations of neurons producing an output which is compared to the original value and their difference is returned into the network. This process adjusts weights until the correct response is learned by administering back propagation. The Radical basis function (RBF) neural net is similar a feed-forward network. It ideally has a quick training time as compared to MLP but may be slower during implementation due to is higher computation power.

Outcomes and Benefits

The generated models were evaluated based on prediction accuracy scores and mean square errors. The prune model took longest to train while RBF model took the shortest training time. The proportion of LOS correctly predicted based on accuracy scores ranging from 77.58% to 78.34%. The mean square error (MSE) for the four models ranged from 0.772 to 0.0813. Thus, based on MSE and predicted accuracy, the RFB model was selected as the best ANN model.