Predicting risk of an adverse event in complex medical data sets using fuzzy ARTMAP network.

Natalya Markuzon[1]*, Stephan A. Gaehde[2]†, Arlene S. Ash[2],
Gail A. Carpenter[1]‡, Mark A. Moskowitz[2]
[1]Department of Cognitive & Neural Systems,
Boston University, 111 Cummington St., Boston, MA 02215
[2]Health Care Research Unit, Section of General Internal Medicine,
Evans Memorial Department of Medicine,
Boston University Medical Center, Boston, MA 02118

Abstract

Fuzzy ARTMAP is a supervised learning system which includes nonlinear dynamics in the learning process. We introduce a new testing procedure which allows the system to estimate the probability of an outcome. Simulations illustrate the system performance in estimating risk in medical procedures. The results are compared to the performance of the logistic regression model. It is shown that both models have similar explanatory power.

Introduction

The application of neural networks to large complex biomedical data sets can improve predictive performance compared to regression modeling. Neural networks can model both linear and nonlinear relationships and update by incorporating emerging patterns. Here we introduce a neural network similar to fuzzy ARTMAP (Carpenter et al., 1992). Fuzzy ARTMAP is a self-organizing pattern classification model for supervised learning and prediction. The performance of the network is compared to that of the logistic regression model, using a cholecystectomy and diabetes data set. The task for each model is to predict the occurrence of an *adverse event* based on information about the patient provided by an input vector. An adverse event is a complication following a medical procedure. Fuzzy ARTMAP was modified to provide probabilistic predictions, as does the logistic regression model. Performance of the two models was similar on these data. This paper includes a short description of the network architecture, the data organization, and the simulation results.

Fuzzy ARTMAP and probabilistic predictions

Fuzzy ARTMAP (Figure 1) includes a pair of Fuzzy ART modules (ART_a and ART_b) (Carpenter, Grossberg and Rosen, 1991) linked together via an inter-ART associative memory F^{ab} that is

*Supported in part by British Petroleum (BP 89A-1204).
†Supported in part by VA Fellowship Training Grant.
‡Supported in part by ARPA (ONR N00014-92-J-4015), the National Science Foundation (NSF IRI-90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).

called a *map field.* During supervised learning, ART_a receives a stream $\{\mathbf{a}^{(p)}\}$ of input patterns and ART_b receives a stream $\{\mathbf{b}^{(p)}\}$ of patterns, where $\mathbf{b}^{(p)}$ is the correct prediction given $\mathbf{a}^{(p)}$. These modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time. The controller is designed to create the minimal number of ART_a recognition categories needed to meet accuracy criteria.

Vigilance parameter ρ_a calibrates the minimum confidence that ART_a must have in a recognition category, or hypothesis, activated by an input $\mathbf{a}^{(p)}$ in order to ART_a to accept that category, rather than search for a better one through an automatically controlled process of hypothesis testing. Lower values of ρ_a enable larger categories to form. These lower ρ_a values lead to a broader generalization and a higher degree of code compression. A predictive failure at ART_b increases ρ_a by the minimum amount needed to trigger hypothesis testing at ART_a, using a mechanism called *match tracking.* Match tracking sacrifices the minimum amount of generalization necessary to correct the predictive error. Hypothesis testing leads to the selection of a new ART_a category, which focuses attention on a new cluster of $\mathbf{a}^{(p)}$ input features that is better able to predict $\mathbf{b}^{(p)}$. Match tracking allows a single ARTMAP system to learn a different prediction for a rare event than for a cloud of similar frequent events in which it is embedded.

The general ARTMAP network learns to classify both binary and analogue vectors. This is accomplished by incorporating fuzzy set-theory operations (Zadeh, 1965) into the dynamics of ART modules which is possible due to close formal similarity between the computations of fuzzy subsethood and ART category choice, resonance, and learning. A normalization procedure called complement coding uses on cells and off cells to represent the input pattern, and preserves the individual feature amplitudes while normalizing the total on cell/off cell vector. Learning is stable because all adaptive weights can only decrease in

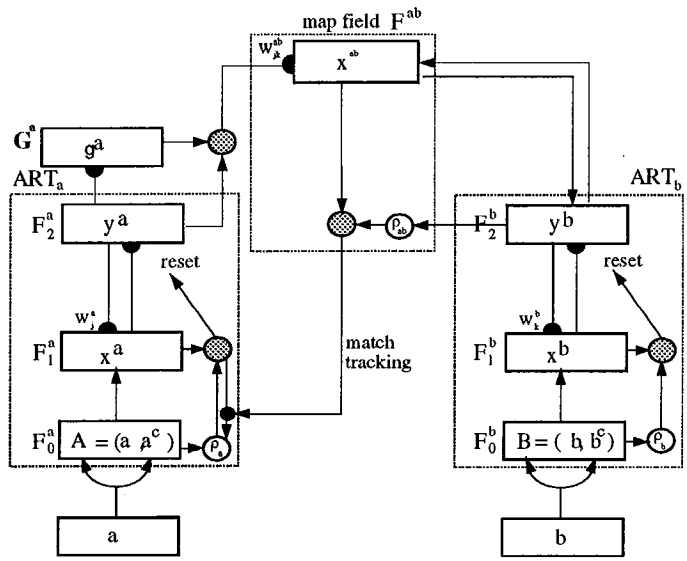

Figure 1: Fuzzy ARTMAP Architecture

time. This feature combined with complement coding and fuzzy logic, lead to increasing sizes of category.

A new testing procedure is introduced which allows the network to, instead of predicting one of many possible outcomes, provide probabilities for each outcome. During testing, instead of selecting just one winner in the ART_a module, K selected winners combine their predictions to yield a probabilistic prediction score. The influence of each category participating in the prediction is weighted by the level of the category's activation, as well as by the total number of training patterns which participated in the formation of the category. Layer G^a stores the number of training inputs for each F_2^a category. Improved prediction is achieved by training the system several times using different orderings of the input set.

An ARTMAP voting strategy is based on the observation that fast learning typically leads to different adaptive weights and recognition categories for different orderings of a given training set, even when the predictive accuracy of all simulations is similar. The different internal category structures cause the set of test set items where errors occur to vary from one simulation to the next. The voting strategy uses an ARTMAP

system that is trained several times on one input set with different orderings. The final prediction for a given test set item is the one made by the largest number of simulations.

Data Description

Two data sets were tested with the network. The first database was a random sample of cholecystectomy cases from 1985 and 1986, selected from the Medicare files of seven states. Abstracted data composed of the presence or absence of about 250 key clinical findings (KCF's) collected using a modified MedisGroups protocol (Iezzoni and Moskowitz, 1988), were provided for each of 3182 patient. Sixteen types of severe adverse events (including mortality within 30 days of admission), which are major complications occurring after cholecystectomy, were defined. The occurrence of any severe adverse event was marked as positive in the binary outcome vector. 16.4% of cases had at least one adverse event. The dimensionality of the input vector was reduced to 16 from the 62 most important KCF's by automatic feature extraction preprocessing. KCF's significantly associated ($p < .05$) with each of the 16 types of adverse events were summed. These 16 sums and an age variable were used to construct an input vector. The models task was to predict the probability of occurrence of a positive outcome after presenting the input vector.

The other database contained data from diabetes patients. We predict an abnormally high evening glucose level based on measurements of patient's blood glucose level and insulin doses during preceding 24 hour period. We employed data about previous evening, morning and the afternoon blood glucose level and regular and NPH insulin doses taken at these times as input information to models. Thus the temporal information about changes in blood glucose level and insulin doses are combined to construct a positionally organized analogue input vector. The outcome, abnormally high versus normal blood glucose level in the evening, is binary. Data from different patients were not combined so that the network was trained individually for each patient. The network was tested on three different patients.

The two data sets described above offer different approaches - in the first one we are predicting the behavior of a new patient based on the experience with previously observed patients, while in the diabetes data we make predictions based on a patient's previous history only.

Results

Both the neural network and the logistic regression model provide a probabilistic estimate of the occurrence of an adverse event while the actual outcome in both data sets is binary. Model performance was evaluated by the C-statistic and R-squared. The C-statistic may be interpreted as the likelihood that the model correctly assigns a higher problem rank when comparing a randomly selected pair of problem/non-problem cases. The C-statistic also is the area under the Receiver Operating Characteristic curve which shows the true-positive rate against the false-positive rate for a given test. R-squared is a measure of the fraction of variation in outcome that is accounted for by a given model. It is an estimate of the explanatory power of a model.

For the cholecystectomy database both models were trained on 4/5 of the data and tested on the remaining 1/5 of the data. Cross validation was performed by randomly dividing the data into 5 equal parts and successively training models to fit each 4/5. The results of the neural network and the logistic regression model were very similar in terms of measures provided (Table 1). C-index for the regression model is equal to 0.68, and $r^2 = 0.065$. The C-index for the fuzzy ARTMAP model is equal to 0.68, and $r^2 = 0.062$. This is a useful level of explanatory power for this task.

regression modeling		ARTMAP modeling	
C-index	R^2	C-index	R^2
0.68	0.065	0.68	0.062

Table 1: Results with the cholecystectomy database

	regression modeling		ARTMAP modeling	
	C-index	R^2	C-index	R^2
patient 1	0.88	0.30	0.91	0.27
patient 2	0.76	0.07	0.71	0.09
patient 3	0.65	0.03	0.73	0.17

Table 2: Results for different patients with the diabetes database

The model performance for different patients in diabetes data set are very different (Table 2). This may partially be explained by physiologic variability among patients. Both the logistic and neural network models were able to successfully predict the elevated blood glucose level based on previously learned variations over the preceding month. The neural network shows superior performance for patient 3, and performance similar to the logistic regression model for patient 1 and patient 2.

We conclude that modified fuzzy ARTMAP is a good alternative for estimating of risk in medical procedures.

References

Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, 759-771.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**, 698-713.

Iezzoni, L.I., and Moskowitz, M.A. (1988). MedisGroups:a clinical assessment. *JAMA*, **260**, p. 3195.

Zadeh, L. (1965). Fuzzy sets. *Information Control*, **8**, 338–353.