THE RECEIVER OPERATING CHARACTERISTIC FUNCTION
AS A TOOL FOR UNCERTAINTY MANAGEMENT IN
ARTIFICIAL NEURAL NETWORK DECISION-MAKING

James M. DeLeo
Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland USA

ABSTRACT

A technique for enhancing artificial neural network (ANN) performance is presented. This technique uses receiver operating characteristic (ROC) methodology to adjust the operating threshold values of ANN output classification processing units to account for both prevalence differences between training cases and real-world cases, and for unequal costs incurred with false positive and false negative classifications.

Keywords: receiver operating characteristic methods, artificial neural networks, back-error propagation, clinical decision-making, uncertainty management.

INTRODUCTION

Artificial neural networks (ANNs) are increasingly used to support highly-skilled, expert decision-making in medicine and many other disciplines. Although users of these biologically inspired computing paradigms claim high success, it is becoming clear that uncertainty issues vexing other computing methodologies could limit ANN performance in many applications. Prevalence and classification error costs are two factors that affect uncertainty. Prevalence is a factor because the frequencies with which various class outcomes occur in ANN training and validation are usually not matched with naturally occurring frequencies in real-world application domains. Classification error costs are a factor since ANN misclassification errors usually incur unequal losses.

This paper proposes the use of ROC methodology as a means to include prevalence and error cost factors in ANN classification strategies for the purpose of enhancing ANN performance. ROC methodology has become well established over the last fifty years as an important tool for addressing decision-making uncertainties in medicine and other disciplines. It has been used to evaluate how well a test or a decision strategy classifies retrospective dichotomous and fuzzy events and to provide a rational basis for designing strategies for classifying similar prospective events [1-4]. ROC methodology easily incorporates prevalence data and misclassification cost information which, in general, are not included in any basic ANN paradigm. Although the technique presented here is illustrated with a back-error propagation neural network, it should be readily extendable to other ANN paradigms, as well as to any other decision-making paradigms that do not consider prevalence and error cost effects where they could significantly affect decision outcomes.

THE BASIC PROBLEM

The basic problem addressed here is illustrated with the back-error propagation ANN shown in Figure 1. Suppose this network has been trained to detect some rare event (e.g. disease, death, etc.) using input data comprised of proven covariate predictive factors for retrospective cases. Now the trained neural network is to be used to decide if the event would be present or absent in individual prospective cases. If the event is determined to be present, then a particular action that is costly financially and/or convenience-wise would be recommended.

Suppose the network has been designed to predict a rare disease and that it has been successfully trained with data representing 1,000 non-diseased cases and 1,000 diseased cases. The first problem is immediately obvious. It has to do with prevalence. The disease is rare, yet the prevalence reflected in the training data is 50%. Does this matter? If so, what should be done about it? The second problem has to do with error costs. Error costs are usually difficult to assess but it is important that they are assessed. The results of such an assessment are usually arrived at after a complex consideration of costs related to trade-offs in accepting false positive and false negative errors. Costs are not only financial but must include qualitative factors such as patient suffering, quality of life effects, etc. For purposes of the technique being developed here, error cost analysis must result in explicitly stating one of the following four values [1]:

1. the acceptable number of true negatives for each true positive.
2. the acceptable number of false negatives for each false
positive.
3. the ratio of false positive cost to false negative cost.
4. the ratio of true negative benefits to true positive benefits.

So the basic task now is to incorporateknowledge of prevalence and error costs when making individual decisions using trained neural networks.

THE BASIC SOLUTION

The approach presented here is based on using ROC functions computed with the sigmoid function values of the trained ANN output classification nodes and the corresponding known classification outcomes for each training case. A basic understanding of receiver operating characteristic (ROC) methodology is offered in another paper by the author appearing in this volume [1]. The companion paper provides methodology for accomplishing the following tasks required in the present work:
1. computing ROC functions.
2. assessing accuracy of a trained neural network by means of the areas under these ROC Functions.
3. calculating slope intercept lines from prevalence and error cost analysis results, determining where these lines touch the convex hull of the ROC Function, and determining new operating points on the sigmoid functions of the output classification nodes that account for prevalence and error costs.

ANN DECISION STRATEGIES

The output classification nodes of back-error propagation ANNs may be thought of as performing fundamental decision strategies. The default strategy implicit in back-error propagation ANN methodology is to assign class 1 if the sigmoid function output value exceeds .5, otherwise to assign class 0. This default strategy does not consider prevalence differences between the training sample and the actual population in which the ANN will be applied. Furthermore it does not account for the costs of making a misclassification.

NEW ANN DECISION STRATEGIES

The companion paper [1] presents the following trichotomous decision rule:

If $x < x_a$ then decide class 0;
If $x > x_a$ then decide class 1;
otherwise, decide “unknown”

in which $x$ is some random variable used for classification purposes, and $x_\alpha$ and $x_\beta$ represent any data point values for $x$ for which $x < x_\alpha$ and $x > x_\alpha$. That paper also presents a procedure for using ROC methodology for selecting appropriate values for $x_\alpha$ and $x_\beta$ to reflect acceptable levels of sensitivity (true positive rate) and specificity (true negative rate) and to incorporate into the selection process the affects of prevalence and error costs. The technique recommended in the present paper is to use the same basic procedure presented in the companion paper to enhance the decision strategies in the classification nodes of the ANN. This technique requires constructing ROC functions using the ANN output node transform function values for each case in the training set as the random variable, $x$, with corresponding known outcome values, $d$.

The companion paper lists the following three elements as required for constructing a ROC function:
1. a test or process that assigns a single value of a test variable, $x$, to each case.
2. a reliable "gold standard" used to definitively classify each case into one or two discrete classes ($d=0$ and $d=1$). Fuzzy classifications are also acceptable.
3. a data set containing $x,d$ data pairs representing various cases.

The output decision nodes of back-error propagation ANNs produce a continuous variable, $x$, as the output of the sigmoid function or any other transform ("squashing") function. The ANN training set provides the "gold standard" value for group classification and the ANN training set provides the $x,d$ training pairs with which to construct the ROC function. The companion paper [1] also shows how to find the point(s) on the ROC to reflect prevalence and error cost issues at which the trichotomous decision rule (1) is constructed. It is suggested here that this process be incorporated after training the neural network and that the trichotomous decision rules are constructed and saved for use during the recall (predictive) use of the ANN.

NEW OPERATING POINT

Adjusting ANN output for prevalence and error costs is done by first finding a single new threshold decision point on the sigmoid function for each output node. Figure 2 illustrates adjustments to the sigmoid function operating point to account for prevalence (left panel) and error cost (right panel) effects. These two illustrations represent graphs of sigmoid functions of back-error propagation ANN processing units which have the following form:
\[
y = \frac{1}{1 + e^{-x}} \tag{2}
\]

in which

\[
x = \sum_{i=0}^{n} w_i a_i \tag{3}
\]

in which \(n\) is the number of node inputs, \(w_i\) is the weight of the \(i\)th input unit and \(a_i\) is the activation level of the \(i\)th input. Note that for \(i=0\), \(a_0=1\), and \(w_0\) represents the bias which controls the left-to-right (horizontal) displacement of the sigmoid function on the \(x\)-axis.

The illustration for the prevalence adjustment (left panel) shows that the default threshold value is at \(x=0\) which corresponds to a prevalence value of \(0.5\), i.e., equal non-diseased and diseased cases. If the actual prevalence in the training set is lower than \(0.5\), then the sigmoid function threshold value corresponding to the lower prevalence lies to the right of \(x=0\) as illustrated with the point marked "training". One way to have the output processing unit operate is to install logic for the trichotomous rule (1) reflecting the new threshold value(s). This approach will also allow the installation of any chosen independent values for sensitivity, and specificity given that the "unknown" state is allowed.

If only a dichotomous rule is required, the easiest thing to do is to shift the sigmoid function to the left so that the "training" point lines up with \(x=0\) and use this shifted sigmoid function for recall classification. This "shift" is easily accomplished by simply adjusting the bias term, \(a_0\) in the exponent of the sigmoid function (3). After adjustment of the operating point is made for training prevalence, it is now possible to adjust the point again by finding a new point on the sigmoid function that accounts for differences in training prevalence and expected real-world application domain prevalence. If prevalence for retrospective cases is lower than for training cases, the point is to the right of the "training" point, thus requiring a left shift of the sigmoid function. If retrospective case prevalence is higher then the training prevalence, then the point is to the left, thus requiring a right shift of the sigmoid function. Both cases are illustrated in the left panel of Figure 2.

Similar adjustments may be made to adjust for classification errors as illustrated in the right panel of Figure 2. Here the adjustments are made from the \(x=0\) default threshold position which corresponds to equal costs for false negative and false positive classifications. Moving the threshold value to the right or shifting the sigmoid function to the left corresponds to higher false negative costs. Moving the threshold value to the left or shifting the sigmoid function to the right corresponds to higher false positive costs.

The companion paper [1] demonstrates how to find one location on the ROC plot that accounts for the combined effects of prevalence and error costs. This corresponds to finding one new point or doing only one shift of the sigmoid function.

**PRACTICAL APPLICATIONS**

Classical engineering ANN applications such as character and speech recognition could obviously be enhanced with prevalence and error cost factors considered since certain characters, radar, and speech patterns occur with greater or lesser frequency than others and more severe costs could occur with a false positive classification, such as in banking, as well as a false negative classification, such as with low-security voice-activated entry systems. Medical ANN decision-making in which single events like relapse or death are predicted, or classifications such as diseases, adverse drug reactions, pregnancy or HIV-positive might also be significantly enhanced by using the technique developed in this paper.

**SUMMARY AND CONCLUSIONS**

A practical technique for addressing prevalence and error cost effects on uncertainty in ANN decision-making has been described. The technique is based on established receiver operating characteristic methodology. It is anticipated that this technique will be applied to appropriate medical applications and that future reports will demonstrate its usefulness.

**ACKNOWLEDGMENTS**

The author gratefully acknowledges Dr. Ramon Tate for his assistance with the graphics and Paul Konopelski for his prompt and careful preparation of the final manuscript.
RARE & COSTLY EVENT PREDICTOR

Figure 1. A back-error propagation ANN for predicting a rare and costly event.

ANN SIGMOID FUNCTION CLASSIFIERS

Figure 2. Illustration of sigmoid function operating threshold adjustments to account for prevalence and error cost factors. The x-axes represent the sum, x, of the weighted input activation levels (3); y-axes represent the output, y, of the sigmoid function (2).

REFERENCES


