DOCUMENT RESUME

ED 372 094 TM 021 756

AUTHOR	Everson, Howard T.; And Others
TITLE	Using Artificial Neural Networks in Educational Research: Some Comparisons with Linear Statistical Models.
PUB DATE	Apr 94
NOTE	19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 5-7, 1994).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; Algebra; *Classification; College Freshmen; Comparative Analysis; *Computer Uses in Education; Discriminant Analysis; *Educational Research; Higher Education; Measurement Techniques; Regression (Statistics); Sex Differences; Statistical Studies; *Student Placement; Test Anxiety
IDENTIFIERS	*Abductory Induction Mechanisms; Linear Models; *Neural Networks

ABSTRACT

This paper explores the feasibility of neural computing methods such as artificial neural networks (ANNs) and abductory induction mechanisms (AIM) for use in educational measurement. ANNs and AIMS methods are contrasted with more traditional statistical techniques, such as multiple regression and discriminant function analyses, for making classification or placement decisions in schools and colleges. Classification rates obtained with multiple regression and discriminant analysis were compared with ANN (back propagation) and AIM methods across a number of plausible models of algebra proficiency that included measures of arithmetic ability, high school achievement, test anxiety, and gender. Analyses were conducted on a sample of 290 male and 310 female college freshmen for the entire sample and for each gender. At each stage 10 randomly selected subsets were used to train and test the neural computing methods. In general, ANN and AIM methods outperformed the more traditional methods. Results suggest that neural computing methods may lead to higher rates of classification accuracy, particularly when underlying models are nonlinear. Included are four tables, and one figure. (Contains 17 references.) (Author/SLD)

Using Artificial Neural Networks in Educational Research: Some Comparisons with Linear Statistical Models[1]

Howard T. Everson[2]
The College Board

David Chance
Department of Psychology
University of Central Oklahoma

Sue Lykins
Department of Psychology
Oklahoma State University

1. This paper was presented at the annual meeting of the American Educational Research Association in New Orleans in April, 1994. This research was supported, in part, by The College Entrance Examination Board.

2. Howard Everson is Senior Research Scientist, The College Board, 45 Columbus Ave., NYC, NY 10023, Tel. 212.713.8301, Internet: htebh@cunyvm.cuny.edu.

Using Artificial Neural Networks in Educational Research: Some Comparisons with Linear Statistical Models

ABSTRACT

This paper explores the feasibility of neural computing methods such as artificial neural networks (ANNs) and abductory induction mechanisms (AIM) for use in educational measurement. We contrasted ANNs and AIM methods with more traditional statistical techniques, such as multiple regression and discriminant function analyses, for making classification and (or) placement decisions in schools and colleges. Our general approach employed a number of plausible classification (i.e., prediction) models of algebra proficiency, using both cognitive and non-cognitive variables. In particular, we compared the classification rates achieved using multiple regression and discriminant analysis with both ANN (back propagation) and AIM methods across a number of plausible models of algebra proficiency that included measures of arithmetic ability, high school achievement, test anxiety and gender. Analyses were conducted for the entire sample, as well as separately for males and females, and at each stage ten randomly selected subsets of the data were used to train and test the neural computing methods. In general, the ANNs and the AIM methods outperformed the more traditional statistical methods, faring better, for example, than linear regression methods when predicting and/or classifying the algebra proficiency of females. These results suggest that neural computing methods may lead to higher rates of classification accuracy, particularly when the underlying models are nonlinear.

Accurate predictions of future academic performance, whether for selection decisions or to make appropriate instructional placements, are central to the college admissions process. As others have noted (Crocker & Algina, 1986; Cronbach, 1971), statistical methods such as linear discriminant function analysis and multiple linear regression are tools widely used to help establish the predictive validity of test scores, and serve as decision aids in the placement and classification process used by schools and colleges. For a variety of reasons, including the essential non-linear nature of many academic achievement models and (or) the complex sets of second and third-order interactions among the predictor variables, these traditional statistical methods do not always yield accurate predictions and (or) classifications. A spate of recent research applying artificial intelligence (AI) computing methods to problems of prediction, selection and classification (see, for example, Lykins & Chance, 1992; Weiss & Kulikowski, 1991) suggests that artificial neural networks (ANNs) and other neural computing methods may substantially improve our classifications, as well as our estimates of the predictive validity of test scores and other educational information.

The purpose of this paper is to explore the utility of neural computing methods to advance research in educational measurement. Neural network computing, unlike conventional computer programming, is a non-algorithmic, non-digital analog, and intensely parallel information processing system. In this study we explore the feasibility of using neural network approach for classifying students as proficient in algebra using a number of cognitive and non-cognitive predictor variables, including prior math achievement, high school gpa, test anxiety and gender. More specifically, we contrast some variations of the back propagation artificial neural network (ANN) and, more briefly, an abductory induction mechanism (AIM) with statistical methods, such as multiple regression and discriminant analysis (Harris, 1975), traditionally used for making classification and (or) placement decisions in schools and colleges.

Limitations of Linear Models

Explaining the relationships among variables is at the heart of the methods used traditionally to establish the predictive validity of educational tests. Common approaches in predictive validity and (or) classification studies include correlation and multivariate regression models. For these models to be useful, however, the relevant variables must be measured with as little error as possible, and the models must fit the data and produce acceptable classification error ratios. It is not uncommon that linear models account for less than half the variance in predictive validity studies (see, for example, Willingham, Lewis, Morgan & Ramist, 1990). When educational tests are used for placement decisions, classification error ratios based on linear models are often unacceptably high, a not so atypical result when the underlying linear model is not a good approximation of the data or, more generally, of academic achievement. Neural computing methods, in contrast, hold promise for developing and testing more complex, nonlinear classification and prediction models with lower classification error ratios than many regression-based approaches, while at the same time achieving reductions in computational complexity. For example, a data set with five predictor variables, using only a second-order regression model, has 20 possible terms with 1,048,575 regressions to perform. Neural computing methods are now being used more widely as alternatives to traditional statistical techniques (Ripley, 1993).

Neural Computing Perspective

Neural computing methods--an outgrowth of artificial intelligence research in the 1950s and 1960s--are a relatively recent development in the information sciences. These methods get their name and biological inspiration because the underlying computational units--networks of processing elements working in parallel--work much like we think neurons functions in the brain (Nelson & Illingworth, 1991). In contrast to most computer programs, neural networks "learn" from a set of exemplary data and are not programmed, as such.

Back propagation networks, for example, are a form of nonlinear regression and are suited for multiple regression applications (Weiss & Kulikowski, 1991). Back propagation networks allow for complex separable classes of information through the use of one or more hidden layers between the input and output layers. The units in a hidden layer can be viewed as a summarizing filter which reduces application dimensionality (Weiss & Kulikowski, 1991). The network assumes that all processing units contribute to the error and therefore propagates the output error backward. The network, in turn, propagates the input forward through the network to the output layer, compares the predicted output to the actual output and determines the amount of error, and then propagates the error back from the output layer to the input layer. This is accomplished by using a gradient descent rule which changes each weight based on the size and direction of the negative gradient on the error surface.

INSERT FIGURE 1 HERE

Since neural network learning procedures are inherently statistical methods, they can be contrasted with ordinary least squares regression techniques. In many cases the ANNs outperform multiple regression techniques in studies involving prediction and classification (Lykins & Chance, 1992).

Both multiple regression and abductive induction mechanisms use a least-mean-square algorithm. Multiple regression techniques partial out the least-mean-square error for each dependent variable and produce a regression function that best fits the data based on the least amount of total error residual. AIM, in contrast, uses a predicted squared error (PSE) criterion to generate a model with the means square error (MSE) intersects with a model complexity penalty value. The AIM approach uses numeric functions which include neural networks as well as higher order functions called abductive networks, thus integrating advanced statistical methods with neural network technology.

In the past few years there have been clear advances in neural computing technologies, and they have now developed to the point where they hold promise for current applications in psychometric research. Neural computing techniques, for example, have been applied with some success in a variety of scientific and engineering settings, including biological research (Weinstein et al., 19[illegible]), economic forecasting (Chance, Cheung, & Fagan, 1992; Shadra & Patil, 1990) and personnel selection and training (Dickieson & Gollub, 1992; Sands, 1991; Sands & Wilkins, 1991). Thus, artificial neural networks and related method appear to hold some promise for educational and psychological research and their potential applications require further exploration.

Design

Since this study was essentially exploratory in nature, no specific sets of hypotheses were tested. Instead, we attempted to define a number of plausible predictive models of proficiency in algebra, using standardized test scores and prior academic achievement variables as well as non-cognitive measures of test anxiety and gender, to compare the accuracy of the predictions and (or) classifications produced by various statistical and neural computing techniques. In general, our models followed the design of conventional predictive validity studies that are widely available in the literature. Two general classes of classification methods were contrasted--general linear models and non-linear, artificial neural network models--and their prediction and classification accuracy rates were compared for both males and females.

METHOD

Our general approach, as we noted earlier, was to contrast a number of plausible classification models of algebra proficiency. In particular, we were interested in understanding the role of both cognitive (i.e., prior academic achievement and math test scores) and non-cognitive (i.e., test anxiety) variables for enhancing our classifications of proficiency in basic algebra under the various statistical and computational models. To achieve this understanding, we compared the classification rates achieved using discriminant function analysis, multiple linear regression, ANN back-propagation and AIM methods across groups of males and females. In all cases the models included a self-report measure of test anxiety (Spielberger, Gonzalez, Taylor, Anton, Algaze, Ross, & Westberry, 1980) , gender, high school grade point average, and a forty standardized multiple choice measure of mathematics, which included two twenty item subscales measuring arithmetic and basic algebra.

Procedures

The participants in this study, 290 men and 310 women, were freshmen at a major urban university. Prior to taking a 40 item standardized multiple-choice mathematics test (which contained two 20 item subscales measuring arithmetic and fundamental algebra), each participant completed a 20 item self-report measure of test anxiety--the *Test Anxiety Inventory* (Spielberger, et al., 1980)--which contains two subscale scores: worry and emotionality. In addition, each completed a questionnaire that asked for demographic information including a self-reported estimate of high school gpa. The forty item mathematics test was administered under timed conditions, and each participant had one hour to complete the test.

A ten-fold cross validation was run to achieve an estimation of the population statistics for the various methods, following Weiss and Kulikowski (1991). The 600 cases were randomly divided into ten training sets (which included 540 cases) and ten test sets (comprised of the remaining 60 cases). Multiple regression and (or) disriminant function analyses were computed on the training sets and a prediction equation was established for each of the corresponding test sets. The correlation between the predicted and the actual algebra proficiency scores was then computed. In addition, ten ANN back propagation and ten AIM networks were run on the training sets and tested on the appropriate test set. Again the correlations between the predicted and actual values of the algebra proficiency scores were computed and contrasted with the correlations derived from the multiple regression analyses.

The algebra proficiency variable was dichotomized for use in the comparisons between the discriminant analysis and the ANNs. A raw score of 10 or below was classified as low (0) and 11 or above as high (1) algebra proficiency. Discriminant analysis classification equations were determined based on the training sets and the correct percentage classified was established for the corresponding test sets. Like our earlier method, training sets were run using neural networks and the percentage of correct classifications were calculated for comparison with the results of the discriminant analyses.

The neural network computer simulations were run using a 486 DX 66 PC. The software package Neural Ware Professional II Plus (1991) was used for all back propagation architectures. Three different variations of back propagation were used, including standard back-propagation, functional links, and extended-delta-bar-delta (see Lykins & Chance for a more complete discussion of these methods). All networks were run with four nodes in a single hidden layer, and weight adjustments were calibrated every thirty epochs. SYSTAT 5.1 (1990) was used for all statistical analyses.

RESULTS

Table 1. presents the means and standard deviations for all variables used in the prediction and (or) classification models.

INSERT TABLE 1 HERE

High school GPA was measured using an ordinal scale; 1= 60-70; 2=71-80; 3=81-90; and 4=91-99. The two components of test anxiety--worry and emotionality--were measured using a scale that requires reporting the frequency of a variety of anxiety symptoms occurring prior to, during, or after an exam. Responses are measured using a four-point Likert-type scale ranging from 1 (almost never) to 4 (always). As noted earlier the TAI yields a score for worry based on a subset of eight items, and a score for emotionality based on another eight-item scale. Table 2 presents the zero-order correlations for this same set of variables.

INSERT TABLE 2 HERE

The results of our comparative analyses of the utility of ANNs versus multiple regression methods are presented in Tables 3A, 3B, and 3C below. Multiple regressions and ANNs performed equally well for all cases, and were well matched for the males in our sample. The ANNs, in general, were unable to detect additional nonlinear information in the data. The ANNs, however, outperformed multiple regression when applied to the data from the females in our sample, suggesting that additional nonlinear information was available in those data.

INSERT TABLES 3A - 3C HERE

Similarly, the results of the contrasts in classification accuracy between the discriminant function analyses and the ANNs are summarized in Tables 4A, 4B, and 4C, respectively.

INSERT TABLES 4A - 4C HERE

In this contrast the ANNs out performed the discriminant analysis method in all three groups. This is not surprising, since neural networks have shown consistently high classification accuracy in other applications (see Lykins & Chance, 1992; Weiss & Kulikowski, 1993). The 15% increase in accuracy for the males clearly demonstrates the viability of using neural networks methods for educational placements and classifications.

Table 5 shows a comparison of multiple regression methods with AIM results for the ten randomizations for all cases.

INSERT TABLE 5 HERE

AIM clearly performed better overall on these data than did the multiple regression method, showing higher correlations for seven of the ten randomizations. When compared with the ANNs, however, the AIM method produced higher correlations in only three of the randomizations (see Table 3A, for example).

Conclusions

We believe this line or research has a number of important outcomes. First, by systematically comparing the traditional linear methods with relatively new neural network models we begin the exploration of ANNs for use in the larger field of educational measurement. Moreover, the number of models generated and tested during the course of comparative research can provide important insights into the validity of test scores and other non-cognitive information for placement and classification decisions. Lastly, this line of research may provide new perspectives into the ways in which many new neural computing methods can be used in conjunction with more traditional statistical approaches to improve our ability to accurately classify and place students into educational experiences that are appropriate for them.

References

Caudill, M., & Butler, C. (1992). *Understanding neural networks, Vol. 1.* MA: Massachusetts Institute of Technology.

Chance, D. , Cheung, J., & Fagan, J. (1992). *Short term forecasting by neural networks while incorporating known characteristics.* Paper submitted for presentation to he International Conference on Neural Networks.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* NY: Holt, Rinehart, and Winston.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (Ed.), *Educational measurement*, 2nd edition. Washington, DC: American Council on Education.

Dickieson, J. , & Gollub, L. (1992). *Artificial neural networks and training.* Independent research and independent exploratory development programs: FY91 annual report. San Diego, CA: Navy Personnel Research and Development Center.

Harris, R.J. (1975). *A primer of multivariate statistics.* New York, NY: Academic Press.

Lykins, S. & Chance, D. (1992). *Comparing artificial neural networks and multiple regression for predictive application.* Paper presented at Southwestern Psychological Society, Austin, TX.

Nelson, M. M. & Illgworth, W.T. (1991). *A practical guide to neural nets.* Reading, MA: Addison-Wesley.

Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA for the predictive validity of the SAT. In W.W. Willingham, C. Lewis, R. Morgan, & L. Ramist (Eds.), *Predicting college grades; An analysis of institutional trends over two decades* (pp. 253-288). Princeton, NJ: Educational Testing Service.

Ripley, B.D. (1993). Statistical aspects of neural networks. In *Chaos and networks--statistical and probabilistic aspects.* Barndorff-Nielsen, O.E., Cox, D.R., Jensen, J.L., and Kendall, W.S. (eds.) London: Chapman & Hall.

Sands, W.A. (1991). *Artificial neural networks: A tool for psychologists.* Paper presented at the 33rd Annual Conference of the Military Testing Association, San Antonio, TX.

Sands, W.A., & Wilkins, C.A. (1991). *Artificial neural networks for personel selection.* Paper presented at the 33rd Annual Confrrence of the Military Testing Association, San Antonio, TX.

Sharda, R., & Patil, R.B. (1990). *Neural networks as forecasting experts: An empirical test.* Proceedings of the IJCNN meeting (pp. 491-494). Washington, DC.

Spielberger, C.D., Gonzalez, H.P., Taylor, C.J., Anton, W.D., Algaze, B., Ross, G.R., & Westberry, L. (1980). *Preliminary professional manual for the test anxiety inventory.* Palo Alto, CA; Consulting Psychologists Press.

Weinstein, J., Kohn, K., Grever, M., Viswanadhan, V., Rubinstein, L. Monks, A. Scudiero, D., Welch, L., Koutsoukos, A. Chiausa, A. & Paull, K. (1992). Neural computing in cancer drug development: Predicting mechanism of action. *Science*, Vol. 258, pp. 447-451.

Weiss, S.M., & Kulikowski, C.A. (1991). *Computer systems that learn: Classification and prediction methods from statistics, neural nets, machine learning, and expert systems.* San Mateo, CA: Morgan Kaufmann.

Willingham, W., Lewis, C., Morgan, R., & Ramist, L. (Eds.) (1990). *Predicting college grades; An analysis of institutional trends over two decades.* Princeton, NJ: Educational Testing Service.

Table 1. Means and Standard Deviations

Variable	N	Mean	SD
HS GPA	600	2.44	0.69
WORRY	600	14.86	4.52
EMOTION	600	16.45	5.22
ARITHMETIC	600	13.24	3.84
ALGEBRA	600	10.45	5.15

Table 2. Zero-Order Correlations

	Sex	HSGPA	Worry	Emotion	Arithmetic	Algebra
Sex	1.0000	.1090**	-.0932*	-.1524**	.2927**	.2943**
HSGPA		1.0000	-.0189	.0038	.2076**	.3187**
Worry			1.0000	.7522**	- .1433**	-.0719
Emotion				1.0000	-.0848*	-.0081
Arith.					1.0000	.6648**
Algebra						1.0000

(Note: * SIG. <.05 and **SIG.<.01)

Tables 3A-3C. Comparisons of Multiple Regression and ANN Methods

3A. All Cases

Test Set	MR	ANN
0	.6287	.6438
1	.6700	.6707
2	.7618	.7698
3	.7048	.6996
4	.6349	.6252
5	.6186	.6260
6	.6832	.6825
7	.7130	.7309
8	.7149	.7167
9	.6858	.6861
mean r	**.6816**	**.6851**

3B. Females

Test Set	MR	ANN
0	.6717	.6972
1	.7230	.7653
2	.5196	.6425
3	.5648	.5718
4	.6239	.6664
5	.7543	.7876
6	.6334	.7437
7	.7275	.7482
8	.7225	.7527
9	.5996	.5915
mean r	**.6540**	**.6894**

3C. Males

Test Set	MR	ANN
0	.6717	.7637
1	.6575	.7674
2	.5706	.6534
3	.6429	.5846
4	.7866	.7935
5	.8037	.7370
6	.6901	.6488
7	.6779	.6229
8	.8363	.8139
9	.8161	.8050
mean r	**.7154**	**.7190**

Tables 4A-4C. Comparisons of Discriminant Analysis and ANN Methods

4A. All Cases

Test Set	DA	ANNs
0	83.3%	86.7%
1	78.3	85.0
2	88.3	85.0
3	78.3	83.3
4	80.0	83.3
5	70.0	70.0
6	73.3	75.0
7	78.3	78.3
8	75.0	76.6
9	78.3	80.0
mean %	78.3	80.3

4B. Females

0	80.0%	86.7%
1	80.8	76.7
2	80.0	83.3
3	70.0	73.3
4	76.6	76.6
5	83.3	86.6
6	60.0	83.3
7	80.0	83.3
8	86.6	83.3
9	69.0	75.9
mean %	76.5	80.9

4C. Males

0	77.0%	74.2%
1	74.2	83.9
2	67.7	67.7
3	61.3	64.5
4	61.3	83.7
5	65.0	93.5
6	54.8	87.1
7	74.2	93.5
8	71.0	93.5
9	83.9	100.0
mean %	69.0	84.2

Table 5. Comparison of Multiple Regression and AIM Methods

Test Set	MR	AIM
0	.6287	.6845
1	.6700	.7324
2	.7618	.5600
3	.7048	.6075
4	.6349	.7661
5	.6186	.6639
6	.6832	.7191
7	.7130	.6553
8	.7149	.8124
9	.6858	.8090
mean r	.6816	.7010

Figure 1. Neural Net Back Propagation Architecture