A Comparison of Three Methods of Matching People to Jobs using Personality

Alan D. Mead

Institute for Personality and Ability Testing

James M. Schuerger

Cleveland State University

Paper presented at the fourteenth annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, Georgia. Please address correspondence to James Schuerger or Alan Mead, PO Box 1188, Champaign, IL, 61824-1188. Or via email, to jkschuerg@aol.com or adm@ipat.com.

Running Head: JOB MATCHING COMPARISON

A Comparison of Three Methods of Matching People to Jobs using Personality

Outplacement and vocational counselors routinely compare personality characteristics of clients with personality characteristics of job incumbents as part of their recommendations to clients. Implied is the expectation that if persons who enter jobs resemble the incumbents, they will be more satisfied, less likely to leave, and more productive than otherwise. While there is some controversy about the effectiveness of this process (e.g., contrast Lowman, 1991 with Spokane, 1991), there is a consensus on the general importance of personality for predicting work satisfaction and performance.

The traditional method of prediction of nominal-level data such as occupational choice is multiple discriminant function analysis (MDFA). Prior research using MDFA (Schuerger & Sfiligoj, 1998) has shown small but consistent differences across studies and hit rates of about 50% (persons correctly classified into occupations). Schuerger (1992) reported hit rates of 25 to 50% using adolescent personality scores to predict adult occupational group by profile similarity coefficient (r_p; Cattell, Eber, & Tatsuoka, 1970). In interest measurement, Strong's early research showed that of men tested in college, somewhat more than 50% ended up in jobs congruent with areas of high interest similarity.

Although hit rates of 50% are fairly good as far as prediction in the social sciences is concerned, there is clearly room for improvement. Possibly, the implicit or explicit assumptions of linear models are not borne out by data (leading to the modest prediction). For example, while the relationship of cognitive ability with important job constructs (e.g., performance) is widely considered linear (Hawk, 1970; Coward & Sackett, 1990), the relationship of personality with job constructs is often assumed to be nonlinear (Brinkmeyer & Rybicki McDaniel, 1998). Thus methods of prediction which make less strong assumptions may produce better predictions.

Relatively new methods of prediction which make fewer assumptions have arisen from machine learning traditions within the artificial intelligence (AI) movement. These methods, once of interest primarily to AI researchers have steadily gained acceptance as valid methods of prediction, especially of nominal level data such as job choice. The two methods of concern in this paper are artificial neural networks (ANNs; Masters, 1993; Ripley & Hjort, 1993) and classification trees (Breiman, Friedman, Olshen, & Stone, 1984; Quinlan, 1993).

Artificial neural networks were first developed by cognitive scientists to model the learning of the human brain. For purely predictive purposes, ANNs are essentially a series of differential equations which are solved in an iterative fashion to predict criteria. In fact, one simple ANN is mathematically equivalent to multiple regression. Features that distinguish ANNs from traditional linear models include a relative paucity of assumptions, an ability to model non-linearity relationships, and a complex and highly iterative parameter estimation. Several recent studies (Collins & Clark, 1993; Santor & Coyne, 1998) used neural networks with great success in some cases while obtaining unimpressive results in others. There is a feeling among ANN enthusiasts that these methods work better in some cases than others. Neural networks have not previously been applied to the prediction of job choice.

Classification trees were developed to infer “IF-THEN-ELSE” relationships (production rules) from data so these could be implemented in a formal logic engine such as an expert system. Rule-sets inferred on the basis of these classification can often present a clear predictive model when presented in an “inverted tree” form (hence the name): Each branching point splits the sample based on some predictive variable with the "leaves" at the bottom indicating group membership. Classification trees have apparently been applied most often to psychometric problems (e.g., Brown & McManus, 1994), were based on CHAID (SPSS, Inc, 1992), and were not encouraging. This research uses the more sophisticated C4.5 method (Quinlan, 1993).

Occupational 16PF (Cattell, Cattell, & Cattell, 1993) personality profiles were available for a large number of individuals representing each of Holland's job types. The purpose of this study was to determine which of these three methods was most effective at predicting job choice from personality profiles. Consistent with the literature on AI prediction methods, we hypothesized that they would fair as well as the MDF approach if the linearity and other assumptions were met and improve upon this traditional approach if its assumptions were not met.

Method

Datasets. Several data sets were available to us which contained 16PF primary scale scores and either Holland codes or occupational information. Some test-takers were applicants when they completed the 16PF questionnaire (only incumbent's data were used in this study) while others took the 16PF questionnaire for developmental or research purposes. About 17% of the sample took the fifth edition of the questionnaire, the reminder responded to a previous edition. About 76% of the data were from the files of organizational practitioners who assessed high performing employees, high-level candidates, valuable employees in need of coaching, and a small number of individuals seeking counseling. About 4% were taken from IPAT scoring and reporting archives. The remainder were participants in field research such as test validation. A total of 4221 individuals were available.

The Holland codes were part of the archived data for the practitioner files. The occupational samples were identified in Gottfredson and Holland (1989). The 4% from IPAT test scoring were self-categorizations made by the test-takers. A subsample of 89 cases with a job title was coded by both authors with an agreement rate of 90% before differences were settled by discussion. The self-categorization of these 89 people agreed with our coding 81% of the time.

Insert Table 1 about here

Procedure. The prediction methods were found to be highly influenced by base rates of the Holland types. To match base rates of the general population as closely as possible, the sample was broken into calibration and holdout based on the common two-thirds rule; two-thirds of the sample were randomly sampled for the calibration sample with the remaining third forming the crossvalidation sample. These samples were then broken into six samples based on Holland code. Cases were sampled randomly from these 12 samples to match the percentage expected in the general population. For Holland codes R, S, & C it was necessary to sample with replacement. The six calibration (crossvalidation) samples were then joined to form a single calibration (crossvalidation) sample with the proper base rates. Table 1 presents the 1980 US Census figures for the US population (quoted from Appendix III or Gottfredson & Holland, 1989), the imbalanced original file percentages, and the final percentages in the calibration and crossvalidation samples.

Results

MFDA results. Table 2 presents the confusion matrices for the MDF analysis of the calibration and crossvalidation samples. A confusion matrix presents an *k* by *k* matrix of actual and predicted classifications for *k* groups. In this case, the actual job types are listed as the row headers and the predicted classifications as the column headers. The entries are the number classified. Thus, a high degree of prediction is shown by large diagonal elements and small off-diagonal elements. In the confusion matrices shown in this paper, the row, column, and overall (in the lower right) percentages are shown. The row percentages show the percent of a given Holland type which is classified correctly. The column percentages show the percent of a given Holland *classification* which is accurate. The overall value shows the percentage in the entire sample classified correctly. Clearly, a matrix could have very good classification rates overall while individual (low base rate) classes were poorly classified.

Insert Table 2 about here

Table 2 shows that the overall classification rates for the calibration and crossvalidation samples were similar (about 56% and 55%) but that the two lowest base rate job types (Investigative and Artistic) were poorly predicted.

ANN results. Unlike linear models such as MDF's, there are a very large number of possible network configurations. Research has shown (see Masters, 1993) that networks with a single hidden layer are able to handle most classification tasks. Even so, the hidden layer may contain any number of nodes. The recommended method of fitting ANNs is to choose a model with few nodes and then increase the nodes until performance degrades (presumably after improving).

Insert Table 3 about here

Table 3 presents the overall accuracy of seven models based on increasing numbers of hidden nodes. While there is some variation, there is not much. We chose the model with six nodes because it predicted accurately in both samples. Undoubtedly this capitalizes on chance but apparently (from Table 3) very little.

Insert Table 4 about here

Table 4 presents the confusion matrix for the ANN. These results are remarkably similar to those in Table 2; the overall level of prediction is a little lower and the shrinkage between calibration and crossvalidation a little larger.

Insert Table 5 about here

Classification tree results. Table 5 presents the confusion matrix for the classification tree. It is marked by a higher degree of accuracy in the calibration sample and a greater degree of shrinkage in the crossvalidation sample. Thus this method stands out as having capitalized on chance to a high degree and having predicted *clearly* worse in the crossvalidation sample. It did classify the Artistic job type better but only slightly.

Discussion

As Table 6 clearly shows, the traditional linear composite method which has been used successfully for decades appears to be the best for the task of predicting Holland job type from personality data. The ANN approach was slightly worse than the MDF approach and the classification tree approach was clearly worse. Our main hypothesis was not borne out by these data.

Insert Table 6 about here

Base rates. Clearly base-rates affected all these procedures. Because we did not vary base rates (e.g., by resampling to match some other mix of job than the 1980 census), we cannot present data on the relative sensitivity of each of these procedures. But that probably does not matter; sensitivity to base rates is *appropriate* and undoubtedly contributes to effective prediction for the group. Nonetheless, in a different population (liberal arts graduates, non-college-bound high school students, etc.) with very different base rates, the models fitted here might perform very differently than is suggested by the crossvalidation confusion matricies.

Criterion reliability. Some reviewers have suggested that the criterion should be a Holland classification based on a test score, such as the SDS (Holland, 1985). This would minimize the effect of people "in the wrong job." We feel that the usefulness of the predicted values would be greater predicting actual job than predicting another test. After all, an interest test could be as easily administered as a personality test.

A separate issue is the reliability of the classification. This seems clear-cut: there was a very high level of agreement between the two authors (90%) and between the test-takers and the authors (81%). The disagreements tended to be over jobs which involved a significant relation to two Holland types, such as a customer service manager or jobs which are broad, such as psychologist. We feel that the Holland codes were accurately assigned.

Caveats and future directions.

It may be that the AI methods were used inappropriately. Quinlan (1993) clearly states that C4.5 was designed to use nominal or perhaps ordinal data and he even discourages the use of C4.5 with continuous data. Nonetheless, if the personality dimensions data had curvilinear relations with job type, C4.5 would have had an advantage. The degree of shrinkage observed suggests that the mechanisms built into C4.5 to prevent shrinkage (primarily "pruning the tree") were insufficient for this type of data. Tweaking the pruning parameters (which were left to their defaults) might well improve the performance of this approach.

Additionally, tweaking these parameters might make the resulting table easier by producing an understandable number of branches. As it was, there were 151 rules produced--far too many to understand.

It might also help the classification tree method to recode the input data to have fewer values. For example, 16PF scales are often interpreted not in their 10-point "native" metric but as high, high-average, low-average, and low. Perhaps data re-written in this form would be more amendable to analysis using classification trees. Possibly classification fit to this recoded data might out-predict MDF's fitted to more detailed, non-recoded data.

The results ANN approach was close to those of the MDF approach so the two methods are similar, right? Wrong! The fitting of the ANN models to produce the results described here took many hours of computer time while the MDF method took only seconds. Although the computations are not infeasibly slow, they will be a clear detraction for anyone who wishes to use these methods. (Classification speed, however, is comparable to that using MDF).

Narrow or specific prediction? Although these results based on seventeen predictors were not supportive of the AI methods, prediction using fewer predictors may yield better results. In other prediction research (Ashton, Jackson, Paunonen, Helemes, & Rothstein, 1995; Mershon & Gorsuch, 1988; Paunonen, 1993) specific, narrow traits consistently outpredicted broad, global traits such as the Big-Five (Goldberg, 1973). However, these studies used linear composites (regression); this effect may not hold for the very different AI methods. The AI methods, which clearly can capitalize on chance, may outperform MDF when the predictors are more reliable, less highly correlated, and fewer in number.

References

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helemes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of research in personality, 29*, 432-442.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman & Hall.

Brinkmeyer, K. R. & Rybicki McDaniel, S. L. (August, 1998). *The dark side of normal personality: New perspectives for workplace decisions*. Paper presented at the 106TH annual convention of the American Psychological Association, San Francisco, CA.

Brown, S., & McManus, M. A. (1994). *Comparison of additive and configural biodata keying methods*. Paper presented at the ninth annual conference for the Society of Industrial and Organizational Psychology, Orlando, FL.

Cattell, R. B., Cattell, A. K., & Cattell, H. E. P. (1993). *Sixteen Personality Factor Questionnaire*. Champaign, IL: IPAT.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, IL: IPAT.

Collins, J. M. & Clark, M. R. (1993). An application of the theory of neural computation to the prediction of workplace behavior: An illustration and assessment of network analysis. *Personnel Psychology, 46*, 503-524.

Coward, W. M. & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology, 75*, 297-300.

Gottfredson, G. D., & Holland, J. L. (1989) *Dictionary of Holland occupational codes*. Odessa, FL: Psychological Assessment Resources.

Hawk, J. (1970). Linearity of the criterion-GATB aptitude relationships. *Measurement and Evaluation in Guidance, 2*, 249-251.

Holland, J. L. (1985). *The Self-Directed Search, Form E*. Odesa, FL: Psychological Assessment Resources.

Lowman, R.L. (1991). *The clinical practice of career assessment*. Washington, D.C.:APA.

Masters, T. (1993). *Practical neural network recipes in C++*. New York: Academic Press.

Mershon, B. & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology, 55*, 675-680.

Paunonen, S. V. (August, 1993). *Sense, nonsense, and the Big Five factors of personality*. Paper presented at the meeting of the American Psychological Association, Toronto.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.

Ripley, N. D. & Hjort, N. L. (1996). *Pattern recognition and neural networks*. London: Cambridge University Press.

Santor, D. A. & Coyne, J. C. (August, 1998). *Applications of neural networks to the classification of diagnostic conditions*. Paper presented at the 106TH annual convention of the American Psychological Association, San Francisco, CA.

Schuerger, J.M. & Sfiligoj, T. (1998). Holland codes and 16PF global factors: Sixty-nine samples. *Psychological Reports*, *82*, 1299-1306.

Schuerger, J.M. (1992). The Sixteen Personality Factor Questionnaire and its junior versions. *Journal of Counseling and Development, 71*, 231-244.

Schuerger, J.M. & Watterson, D.G. (1996). *Occupational Patterns in the 16 Personality Factor Questionnaire*. Chagrin Falls, OH: WAI, Inc.

SPSS, Inc. (1992). SPSS/PC+ CHAID [Computer program]. Chicago: Author.

Spokane, A.R. (1991). *Career Intervention*. Prentice Hall.

Table 1. Holland code base rates of population, original sample, and resampled samples.

Holland Type	1980 Census	Original	Calibration	Holdout
Realistic	33.2%	22.7%	33.1%	32.9%
Investigative	3.9	4.4	4.3	4.5
Artistic	1.4	1.6	1.5	1.8
Social	13.7	12.2	13.7	13.6
Enterprising	21.8	44.3	21.7	21.6
Conventional	25.9	14.5	25.8	25.7
N	n/a	4221	2813	1413

Table 2. Discriminant function analysis confusion matrices.

Calibration Sample

	Predicted						
Actual	R	I	A	S	E	C	%
Realistic	692	1	1	32	105	79	76.0
Investigative	44	5	0	4	41	26	4.2
Artistic	7	0	4	6	11	13	9.8
Social	93	0	2	154	56	79	40.1
Enterprising	177	3	2	30	256	140	42.1
Conventional	122	1	0	44	107	446	61.9
%	61.0	0.0	0.0	57.0	44.4	57.0	55.9

Crossvalidation Sample

	Predicted						
Actual	R	I	A	S	E	C	%
Realistic	334	3	0	14	59	51	72.5
Investigative	22	0	0	5	17	19	0.0
Artistic	5	0	0	4	6	10	0.0
Social	53	0	3	71	19	43	37.6
Enterprising	86	2	0	22	131	64	43.0
Conventional	46	1	1	25	47	243	66.9
%	61.2	0.0	0.0	50.4	47.0	56.5	55.4

Table 3. Results of the neural network prediction across seven models.

Number of Hidden Nodes	Calibration	Crossvalidation
2	52.2	50.1
3	53.5	51.1
4	59.2	49.1
5	55.6	50.2
6	57.7	50.5
7	58.1	50.4
8	56.6	51.0

Table 4. Confusion matrices for the neural network model with six hidden nodes.

Calibration Sample

	Predicted						
Actual	R	I	A	S	E	C	%
Realistic	692	0	0	18	120	80	76.0
Investigative	43	0	0	5	41	31	0.0
Artistic	9	0	0	5	7	20	0.0
Social	99	0	0	155	64	66	40.4
Enterprising	159	0	0	23	282	144	46.4
Conventional	99	0	0	35	108	478	66.4
%	62.9	0	0	64.3	45.3	58.4	57.7

Crossvalidation Sample

	Predicted						
Actual	R	I	A	S	E	C	%
Realistic	314	0	0	18	89	40	68.1
Investigative	21	0	0	1	16	25	0.0
Artistic	5	0	0	3	5	12	0.0
Social	46	0	0	60	25	58	31.7
Enterprising	94	0	0	12	127	72	41.6
Conventional	55	0	0	45	54	209	57.6
%	58.7	0	0	43.2	40.2	50.2	50.5

Table 5. Classification tree confusion matrices.

Calibration Sample

	Predicted						
Actual	R	I	A	S	E	C	%
Realistic	711	0	0	1	210	8	76.5
Investigative	10	6	0	1	97	7	5.0
Artistic	1	0	9	2	30	0	21.4
Social	11	1	0	256	109	7	66.7
Enterprising	21	1	1	4	561	22	92.0
Conventional	7	0	0	4	174	541	74.5
%	93.4	75.0	90.0	95.5	47.5	92.5	74.1

Crossvalidation Sample

	Predicted						
Actual	R	I	A	S	E	C	%
Realistic	204	0	0	26	192	43	43.9
Investigative	7	0	1	4	40	11	.0
Artistic	4	1	1	4	11	4	4.0
Social	28	3	1	44	96	20	22.9
Enterprising	44	0	2	26	183	50	60.0
Conventional	24	0	3	43	186	107	29.5
%	65.6	0.0	12.5	29.9	25.8	45.5	38.1

Table 6. Comparison of the percentage of correct classifications for the three methods.

Sample	Discriminant Analysis	Classification Tree	Neural Network
Calibration	55.9	74.1	57.7
Crossvalidation	55.4	38.1	50.5