*Proceedings of the 2004 Systems and Information Engineering Design Symposium*
*Matthew H. Jones, Stephen D. Patek, and Barbara E. Tawney, eds.*

# LEARNING FROM STUDENT DATA

Kash Barker, Theodore Trafalis, and Teri Reed Rhoads

School of Industrial Engineering
University of Oklahoma
Norman, Oklahoma

## ABSTRACT

An abundance of information is contained on every college campus. Many academic, demographic, and attitudinal variables are gathered for every student who steps on campus. Despite all this information, colleges still struggle with graduation rates. This is an apt example of an overload of information but a starvation of knowledge. This paper introduces the use of neural networks and support vector machines, both nonlinear discriminant methods, for classifying student graduation behavior from several academic, demographic, and attitudinal variables maintained about students at the University of Oklahoma.

## 1 INTRODUCTION

The University of Oklahoma, echoing statements heard around the higher education community, has recently mandated that increasing graduation rates is a top priority. The attrition of students at the university level is a serious issue and is now being treated as such. Most students leave school in the first two years of college (Tinto 1987). Nationally, only 55 percent of students that began college in the fall of 1995 attained a degree from that college within six years (Wirt 2003). At the University of Oklahoma, 51 percent of freshmen in the fall of 1995 earned an OU undergraduate degree within six years, a common statistic used to compare graduation success rates.

The ultimate goal of higher education is to produce graduates, and several factors can be considered as motivation. Universities face a loss in revenue for each student who does not complete their degree. From a recent study by Valdosta State University, part of the University System of Georgia, ten students who do not persist past their first semester cost the university $326,811 in lost revenues (Valdosta State University 2000). This is a significant amount when considering that roughly 20 percent of students entering college nationally in the fall of 1999 did not return to their respective college for a second year (Consortium 2002). Governments are increasingly calling for higher education to account for the money that is invested in institutions (Yorke 1999). Additionally, fewer graduations lead to fewer alumni and fewer alumni gifts. A major accolade universities strive to achieve is a favorable ranking in *US News* college rankings. Retention and graduation rates make up a considerable portion of the *US News* score, 20 percent and 5 percent respectively (Morse *et al* 2004). With the national image of the university coming in large part from this publication, improving graduation is of concern. Finally, in a time when states vie for corporate expansion, companies often consider the number of college graduates a state can offer in choosing where to expand. Oklahoma, according to the 2000 Census, falls over 6 percent below the national average with 20.4 percent of its citizens holding bachelor's degrees or higher (U.S. Census Bureau 2003).

### 1.1 Problem Statement

Technological developments have allowed for better collection and, more importantly, analysis of large amounts of data. The need for understanding large data sets is found in engineering, business, medicine, and many other areas (Jain *et al* 2000). Useful knowledge is hidden in many of these data sets, and there is an increased need to discover and use this knowledge, particularly in student data. The knowledge of interest to a university could include knowing whether a student is likely to graduate in, say, six years.

Classifying students based on pre-enrollment information would allow the university to identify students who would be "at risk" of failure before they step into the classroom. Support systems, such as orientation, advising, and mentoring programs, could be put into place to positively impact the academic successes of such students.

### 1.2 Pattern Classification Background

Pattern classification deals with learning from a number of inputs to extract an output. In pattern classification, these inputs each belong to one of a number of predetermined output classes. Given a set of training samples belonging to known classes, the problem of classification deals with finding a generalized classifier to predict the class for any new sample.

A number of inputs, or attributes, and their corresponding outputs, or classes, are given. A training algorithm uses these sample inputs, called the training set, to design a decision function that can accurately predict the class for any sample thereafter. Given are $n$ training samples ($\mathbf{x}_i$, $t_i$), for $i = 1, ..., n$, where $\mathbf{x}_i \in R^d$ is a vector of $d$ attributes for the sample and $t_i \in \{-1, 1\}$ represents the corresponding class for the sample. For the problem at hand, a student can fall into one of only two classes: the student obtained an undergraduate degree at the University of Oklahoma within six years or that student failed to do so. A testing set of known attributes and response values is then established. The algorithm response is compared to the actual response to determine how well the classifier performs.

## 2 PREVIOUS EDUCATIONAL RESEARCH

Investigations in educational research, particularly in the area of retention and graduation of college students, have taken place for many years. Popular theoretical models of student attrition proposed by Spady (1971) and Tinto (1987) date back several decades. However, data mining techniques, while proposed many years ago, have only been researched recently, and new applications have appeared frequently of late.

### 2.1 Variables in Academic Research

Demographic variables such as gender (Tinto 1987, Hermanowicz 2003), race (Tinto 1987, Hermanowicz 2003), hometown population (Aylesworth *et al* 1976), and finances (Yorke 1999, McDaniel *et al* 2001) have been established as predictors of college graduation in previous research. Some academic variables previously proposed include high school gpa and college entrance exam scores (McDaniel *et al* 2001, Hermanowicz 2003).

Several attitudinal factors that affect college graduation have been cited: extracurricular involvement (Tinto 1987), happiness with the social environment and choice of major (Yorke 1999), hours of study per week and the desire to earn a bachelor's degree (McDaniel *et al* 2001), and nonintellectual factors including examination anxiety and fear of failure (Larose *et al* 1998).

### 2.2 Modeling in Educational Research

Several modeling techniques have been applied previously to student data: logistic regression (Dey *et al* 1993), structural equation modeling (Cabrera *et al* 1993), Markov processes (Heiberger 1993), and proportional hazards modeling (Murtaugh *et al* 1999).

Very few instances of nonlinear discriminant methods are found in educational research. Byers-Gonzalez and DesJardins (2002) used neural network techniques to predict college application behavior at a public university in Iowa and compared the results with traditional statistical methods used to study student application behavior. Walczak (1994) developed a neural network system to determine whether a student would actually attend a particular university if accepted.

## 3 DATA

Upon entrance to the University of Oklahoma, freshman students are placed in University College regardless of their major. The primary role of University College is to advise new students on appropriate courses, selection of major, and OU resources and policies. Much effort is expended by University College to keep students at the University past their freshman year.

University College administers a survey to all incoming freshman upon enrollment to OU and has been doing so for several years. On the survey, which varies from year to year, are several questions of attitudinal and demographic nature. Many of the questions found on the survey closely resemble student attributes appearing in the educational research previously discussed. Table 1 displays a few of the roughly 100 questions asked on the surveys. In addition to survey results, further academic and demographic information was gathered. This information, such as state of residence, standardized test scores, and ethnicity, is maintained by the University's Information Technology department.

Table 1. Survey question examples

| Question | Response |
|---|---|
| "Members of my immediately family attended OU." | 1. yes, 2. no |
| "I expect to pledge a sorority or fraternity." | 1. yes, 2. no, 3. uncertain |
| "[in high school], I had difficulty motivating myself to study or attend class." | 1. strongly disagree, ..., 10. strongly agree |
| "[at OU], I feel that I will fit in well on campus." | 1. strongly disagree, ..., 10. strongly agree |
| "I need to work to afford to go to school." | 1. strongly agree, ..., 5. strongly disagree |

Records were obtained for freshmen enrolling at OU for the Fall 1995, Fall 1996, and Fall 1997 semesters, so chosen because these represent the three most recent semesters for which six-year graduation rates of first-time, full-time freshmen can be calculated.

In all, the student attributes studied included 56 demographic, academic, and attitudinal variables.

Table 2 compares the graduation performance of the three cohorts. The distributions of all 56 variables were examined for differences in student attributes among the cohorts. Generally, all three cohorts behaved similarly, though a more stringent admission policy came into effect before the 1996 cohort students were admitted. Also, OU has seen an increase in the quality of incoming freshmen for several years.

Table 2. Complete data set comparison by cohort

| Cohort | Complete Information | |
|---|---|---|
| | **Count** | **6-year Grad Rate** |
| 1995 | 1774 | 47.4% |
| 1996 | 1486 | 45.8% |
| 1997 | 1833 | 48.3% |

While previous work on estimating missing values was examined and considered (Liu *et al* 1997), only students with the entire set of 56 attributes were studied. The graduation rates of students with incomplete information are considerably different than those students for whom all 56 variables exist, as shown in Table 3. With the difference in graduation rates, one could conclude that the students with incomplete information possess different characteristics that those with complete information. It is regrettable that eliminating some students may not fully capture the entire student population, but for the analysis methods, no missing data can be allowed.

Table 3. Incomplete data set comparison by cohort

| Cohort | Incomplete Information | |
|---|---|---|
| | **Count** | **6-year Grad Rate** |
| 1995 | 744 | 35.8% |
| 1996 | 869 | 40.3% |
| 1997 | 729 | 32.5% |

Each student in the complete data set falls into one of two classes: those who do not graduate within six years and those who do graduate within six years. The freshman major of the student or number of majors the student had over his/her career did not affect the class into which the student fell from a data collection perspective. The class depends only on whether the student received a diploma within six academic years of his/her freshman year.

## 4 METHODOLOGY

In many practical cases, the training data set cannot be separated with a linear classifier. Two nonlinear discriminant methods were used in classifying student graduation: neural networks and support vector machines.

### 4.1 Neural Networks

Tsoukalas and Uhrig (1997) define a neural network as: "A data processing system consisting of a large number of simple, highly interconnected processing elements (artificial neurons) in an architecture inspired by the structure of the cerebral cortex of the brain."

Figure 1 details the general architecture of a neural network. An input layer of $d$ nodes corresponds to the $d$ attributes exhibited by the student. The hidden layer(s) consists of several predetermined hidden nodes connected to the input layer with a set of weights. The hidden layer is then connected to the output layer, for which there is one node due to the nature of the student classification problem, also via a set of weights. The backpropagation algorithm updates these weights using the difference of the actual response and the function response for a given epoch. The neural network algorithm runs for a certain number of epochs or until a minimum error is calculated. Following the completion of the algorithm, the weights are then used to classify further data.

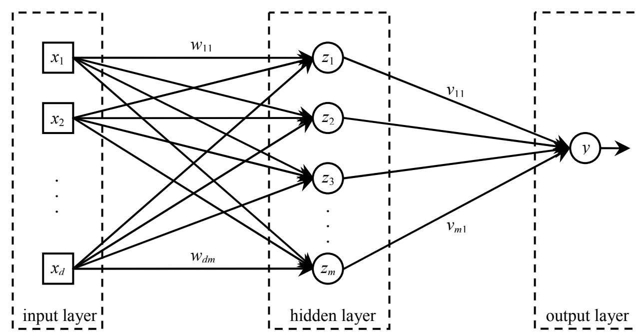


Figure 1. Neural network architecture

For the student classification problem, the number of epochs was held constant at 1000, the tangent sigmoid activation function was always used due to the nature of the desired response, and the training algorithm parameters were held constant. The number of hidden layers and hidden nodes was varied to find the most generalized set of weights.

### 4.2 Support Vector Machines

Support vector machines (SVM), attributed to Vapnik (1995), is a relatively new approach for pattern classification. The performance of SVM rivals or exceeds that of competing methods, including neural networks.

The purpose of support vector machines in classification is to find a separating hyperplane able to separate two sets of data by maximizing the distance, or margin, between the two data sets. This margin is then divided by the separating hyperplane. The theory is that the hyperplane should be found in the maximum margin so as to optimally separate the two classes. A larger margin corresponds to better generalization (Cherkassky *et al* 1998). To find the maximum margin and separating hyperplane, we solve a convex optimization problem which, theoretically, gives an optimal solution unlike the neural network algorithm that minimizes a non-convex error function.

Figure 2 graphically depicts a linearly separable data set of two classes. The margin between the two classes is shown to be $2/\|\mathbf{w}\|$, where **w** is the vector of weights that is used in the classifier. The darkened observations are the support vectors, or the outer observations of each class through which the bounding hyperplanes lie. These support vectors are very important, as they are only points necessary in separating the two classes. The margin between the two classes is maximized for better generalization using quadratic optimization or equivalently minimizing the norm of **w**.

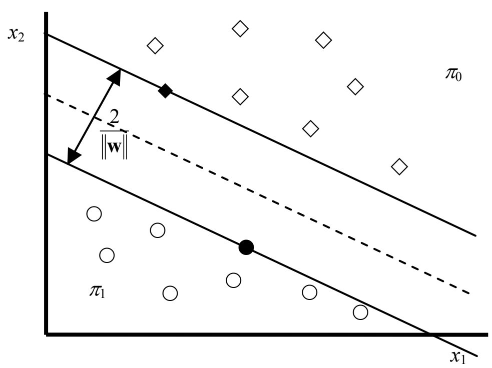


Figure 2. Depiction of margin for a linearly separable example

For a nonlinearly separable set of data, much like the problem of student data, observations can be mapped to a higher dimensional space through the use of a nonlinear kernel function. Common kernel functions, ones used in this research, appear in Table 4.

Table 4. Common kernel functions

| Kernel | Functional Form |
|---|---|
| Linear | $K(x,x') = \langle x,x' \rangle$ |
| Polynomial | $K(x,x') = (\langle x,x' \rangle + \mu)^p$ |
| Radial Basis Function | $K(x,x') = \exp\left(-\gamma \|x-x'\|^2\right)$ |

Depending on the choice of kernel, there are several parameters to adjust for improved performance. There is also a penalty $C$ in the constraint functions of the optimization program which can vary for results. The value of C suggested by Cherkassky and Ma (2002) is found with $C = \max\left(\left|\bar{y} + 3\sigma_y\right|, \left|\bar{y} - 3\sigma_y\right|\right)$. Based on preliminary experiments of several values of $C$, it was determined that $C = 3$, found from the above equation, performed better than other values using cross-validation techniques. This value of $C$ was held constant for this research.

## 5 ANALYSIS

The data were partitioned in several different ways for analysis. After partitioning, each variable was normalized to have a mean of zero and standard deviation of one, as such preprocessing techniques are used to reduce the range of each variable.

### 5.1 Combined Data

The first method of analysis involved the data for all three cohorts being combined into a larger pool of 5093 students with no identification given to the cohort to which the student belonged. From that larger pool of students, training sets consisting of 3500 students, or roughly 70 percent, and testing sets consisting of 1593, or roughly 30 percent, were created.

Ten different training and corresponding testing sets were created randomly. The algorithms were performed on all ten sets to better generalize results.

### 5.2 Training/Testing Between Years

The second method for analyzing the data dealt with training with all the students in a given year and testing with all the students in the following year. There are two combinations of interest here: training the algorithm with the 1995 cohort and testing with the 1996 cohort and training the algorithm with the 1996 cohort and testing with the 1997 cohort. The data sets used for each cohort were given previously in Table 2.

### 5.3 Training/Testing Among Years

The third method studied was training and testing within each cohort to determine if a particular cohort had certain properties that allowed for good predictive results when trained with students only in that cohort. Training sets for each cohort consisted of approximately 70 percent of the cohort students, with the remaining 30 percent falling into the testing set. These training and testing sets were randomly produced five times to look at the overall average of algorithm performance.

For example, the 1774 students of the 1995 cohort were divided into a training set of 1200 and a testing set of 574. Table 5 shows the number of students making up the training and testing sets for the three cohorts.

Table 5. "Among years" training/testing set sizes

| Cohort | Training Set | Testing Set |
|---|---|---|
| 1995 | 1200 | 574 |
| 1996 | 1050 | 436 |
| 1997 | 1300 | 533 |

It was presumed that the "combined data" analysis method would provide the best results, or lower percentage of misclassified students. The "between years" analysis was used to determine if it was possible to predict the performance of students from a given year based on students entering a year prior. Finally, the "among years" analysis method was used to test if any particular cohort had characteristics that would allow for classification of its own students but not students of another cohort.

## 6 RESULTS

A synopsis of the numerical results are presented below in the same manner as the data sets were discussed in Section 5. Each of the following tables presents the lowest percentage of misclassified students given the particular algorithm parameters along with the percentage misclassified when testing the algorithm with the set used to train the algorithm. Testing with the training set helps to determine if the large amount of misclassified students is due to the algorithm or the nature of the data.

### 6.1 Combined Data

The results of the "combined data" analysis are shown in Table 6.

Table 6. "Combined data" results

| *Neural Networks* | | |
|---|---|---|
| **Network Architecture** | **Percent Misclassified** | **Testing with Training Set** |
| [1 1] | 36.6% | 32.5% |
| *Support Vector Machines* | | |
| **Kernel Type** | **Percent Misclassified** | **Testing with Training Set** |
| Linear | 36.6% | 34.7% |
| Polynomial ($p$=1) | 36.6% | 34.7% |
| Radial Basis ($\gamma$=0.001) | 36.6% | 33.9% |

Both discriminant methods with architecture and kernel parameters produced the same best-case misclassification percentage of 36.6%.

### 6.2 Training/Testing Between Years

Table 7 shows the results of the analysis between years.

Table 7. "Training/testing between years" results

| *Neural Networks* | | |
|---|---|---|
| **Network Architecture** | **Training/Testing Sets** | **Percent Misclassified** |
| [1 1 1] | 1995/1996 | 39.6% |
| | 1996/1997 | 38.4% |
| *Support Vector Machines* | | |
| **Kernel Type** | **Training/Testing Sets** | **Percent Misclassified** |
| Linear | 1995/1996 | 38.7% |
| | 1996/1997 | 37.9% |
| Polynomial ($p$=1) | 1995/1996 | 38.8% |
| | 1996/1997 | 38.0% |
| Radial Basis ($\gamma$=0.001) | 1995/1996 | 37.7% |
| | 1996/1997 | 37.9% |

The support vector machine mapped using a radial basis function kernel with parameter $\gamma$ = 0.001 produced the lowest percentage of misclassified students. Table 8 shows the percentage of training set students misclassified when testing is performed on the testing set.

Table 8. "Training/testing between years" results

| *Neural Networks* | | |
|---|---|---|
| **Network Architecture** | **Training/Testing Sets** | **Percent Misclassified** |
| [1 1 1] | 1995/1995 | 37.0% |
| | 1996/1996 | 38.8% |
| | 1997/1997 | 35.6% |
| *Support Vector Machines* | | |
| **Kernel Type** | **Training/Testing Sets** | **Percent Misclassified** |
| Linear | 1995/1995 | 34.0% |
| | 1996/1996 | 34.9% |
| | 1997/1997 | 33.3% |
| Polynomial ($p$=1) | 1995/1995 | 34.2% |
| | 1996/1996 | 34.9% |
| | 1997/1997 | 33.3% |
| Radial Basis ($\gamma$=0.001) | 1995/1995 | 33.3% |
| | 1996/1996 | 34.9% |
| | 1997/1997 | 32.1% |

### 6.3 Training/Testing Among Years

The results of the analysis among cohort years are found in Tables 9, 10, and 11.

Table 9. 1995 "Training/testing among years" results

| *Neural Networks* | | |
|---|---|---|
| **Network Architecture** | **Percent Misclassified** | **Testing with Training Set** |
| [1 1] | 39.2% | 23.8% |
| *Support Vector Machines* | | |
| **Kernel Type** | **Percent Misclassified** | **Testing with Training Set** |
| Linear | 38.4% | 33.2% |
| Polynomial ($p$=1) | 38.4% | 22.4% |
| Radial Basis ($\gamma$=0.0005) | 37.9% | 22.8% |

Table 10. 1996 "Training/testing among years" results

| *Neural Networks* | | |
|---|---|---|
| **Network Architecture** | **Percent Misclassified** | **Testing with Training Set** |
| [1 1 1] | 40.2% | 24.5% |
| *Support Vector Machines* | | |
| **Kernel Type** | **Percent Misclassified** | **Testing with Training Set** |
| Linear | 40.1% | 32.6% |
| Polynomial ($p$=1) | 40.1% | 22.9% |
| Radial Basis ($\gamma$=0.001) | 38.2% | 31.3% |

Table 11. 1997 "Training/testing among years" results

| *Neural Networks* | | |
|---|---|---|
| **Network Architecture** | **Percent Misclassified** | **Testing with Training Set** |
| [1 1] | 36.9% | 25.2% |
| *Support Vector Machines* | | |
| **Kernel Type** | **Percent Misclassified** | **Testing with Training Set** |
| Linear | 35.9% | 32.7% |
| Polynomial ($p$=1) | 36.1% | 23.2% |
| Radial Basis ($\gamma$=0.001) | 34.6% | 28.0% |

Generally when training and testing within each cohort, performance was best when using the radial basis function kernel.

### 6.4 General Results

For the "combined" and "between years" data sets, results of testing the algorithms with the training set appear to be somewhat close to the general testing results. When examining the "among years" results, in a much lower misclassified percentage when testing with training set.

Specificity, or the probability of predicting a student will not graduate when that student did not graduate, is roughly 0.7 for all data sets.

Notice that the best performing polynomial kernel always occurred when the exponent $p$ = 1. This closely resembles the linear kernel, and, therefore, the linear kernel and polynomial kernel results are generally the same.

It is worth noting that principal component analysis was also used to reduce the number of variables from 56. Eliminating those variables contributing less than 2 percent of the overall variability, 14 variables remained. However, analysis with these reduced data sets produced results that were worse, sometimes much worse, than the results from the normalized data sets of 56 variables.

## 7 CONCLUSIONS

It appears that training the algorithms with a set of data made of up students from several years results in a more generalized result, though all three methods of partitioning the data differed very little.

Validating the algorithm by testing with the training data often led to similar results when compared to the general testing results. This would lead to the conclusion that we may lack domain knowledge about the data set, and that poor results are not the fault of the algorithm. It may be impossible to capture sufficient and proper knowledge about student graduation.

The percentage of misclassified students is pretty high, as approximately one-third are misclassified. While more work needs to be performed to improve these results, it is important to note the purpose of the study. It is desired to identify students that may need particular attention. While a misclassification error close to zero is desired, the intention of the study may not require that.

## ACKNOWLEDGMENTS

We are grateful to Dr. Myrna Carney of University College and the Information Technology department at the University of Oklahoma for providing us with data.

## REFERENCES

Aylesworth, L.S. and B.L. Bloom. 1976. College experiences and problems of rural and small-town students. *Journal of College Student Personnel* 17: 236-242.

Byers-Gonzalez, J.M., and S.L. DesJardins. 2002. Artificial neural networks: A new approach to predicting application behavior. *Research in Higher Education* 43 (2): 235-258.

Cabrera, A.F., A. Nora, and M.B. Castaneda. 1993. College persistence: structural equations modeling test of an integrated model of student retention. *Journal of Higher Education* 64: 123-139.

Cherkassky, V. and F. Mulier. 1998. *Learning from Data: Concepts, Theory, and Methods*. New York: Wiley and Sons.

Cherkassky, V. and Y. Ma. 2002. Selecting the loss function for robust linear regression. under review in *Neural Computation*.

Dey, E.L. and A.W. Astin. 1993. Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in Higher Education* 34 (5): 569-581.

Hermanowicz, J.C. 2003. *College attrition at American research universities: Comparative case studies*. New York: Agathon Press.

Heiberger, R.M. 1993. Predicting next year's enrollment: Survival analysis of university student enrollment histories. *Proceedings of the American Statistical Association*, *Social Statistical Section*: 143-148.

Jain, A.K., R.P.W. Duin, and J. Mao. 2000. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (1): 4-37.

Larose, S., D.U. Robertson, R. Roy, and F. Legault. 1998. Nonintellectual learning factors as determinants for success in college. *Research in Higher Education* 39 (3): 275-297.

Liu, W.Z., A.P. White, S.G. Thompson, and M.A. Bramer. 1997. Techniques for dealing with missing values in classification. *Advances in Intelligent Data Analysis: 2nd International Symposium*: 527-536.

McDaniel, C., and S.W. Graham. 2001. Student retention in an historically black institution. *College Student Journal* 35: 143-157.

Morse, R.J. and S.M. Flanigan. 2004. Using the rankings. Available online via <http://www.usnews.com/usnews/edu/college/rankings/about/04rank_brief.php> [accessed March 24, 2004].

Murtaugh, P.A., L.D. Burns and J. Schuster. 1999. Predicting the retention of university students. *Research in Higher Education* 40 (3): 355-371.

Spady, W.G. 1971. Dropouts from higher education: Toward an empirical model. *Interchange* 2 (3): 38-62.

The Consortium for Student Retention Data Exchange. 2002. First-year retention rates. Available online via <http://tel.occe.ou.edu/csrde/retnrate2.htm> [accessed March 14, 2004].

Tinto, V. 1987. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago: University of Chicago Press.

Tsoukalas, L.H. and R.E. Uhrig. 1997. *Fuzzy and Neural Approaches in Engineering*. New York: Wiley and Sons.

U.S. Census Bureau. 2003. Educational attainment in the United States: March 2002. <http://www.census.gov/population/www/socdemo/education/ppl-169.html> [accessed March 14, 2004].

Valdosta State University. 2000. Task force on student retention committee report. <http://www.valdosta.edu/retention/report.htm> [accessed on March 24, 2004].

Vapnik, V. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.

Walczak, S. 1994. Categorizing university student applicants with neural networks. *Proceedings of IEEE International Conference on Neural Networks:* 3680-3685.

Wirt, J., S. Choy, S. Provasnik, P. Rooney, A. Sen, and R. Tobin. 2003. *The Condition of Education 2003*. National Center for Education Statistics.

Yorke, M. 1999. *Leaving Early: Undergraduate Non-completion in Higher Education*, Philadelphia: Falmer Press.

## AUTHOR BIOGRAPHIES

**KASH BARKER** is a graduate student in the School of Industrial Engineering at the University of Oklahoma. He received his BS in Industrial Engineering from the University of Oklahoma. His research interests include the application of modeling and data mining techniques. He is a member of NSPE, IIE, SIAM, INFORMS, and American Mathematical Society. He can be contacted by e-mail at <enchilada@ou.edu>.

**THEODORE B. TRAFALIS**, PhD, is a Professor in the School of Industrial Engineering at the University of Oklahoma. Dr. Trafalis earned his BS in Mathematics from the University of Athens, Greece, his MS in Applied Mathematics, MSIE, and his PhD in Operations Research from Purdue University. He is a member of ORSA, SIAM, Hellenic Operational Society, International Society of Multiple Criteria Decision Making, and the International Society of Neural Networks. He is listed in the 1993/1994 edition of Who's Who in the World. He was a visiting Assistant Professor at Purdue University (1989-1990), an invited Research Fellow at Delft University of Technology, Netherlands (1996), and a visiting Associate Professor at Blaise Pascal University, France and at the Technical University of Crete (1998). He was also an invited visiting Associate Professor at Akita Prefectural University, Japan (2001). He can be contacted by e-mail at <ttrafalis@ou.edu>.

**TERI REED RHOADS**, PhD, is Director of Engineering Education for the College of Engineering and Assistant Professor in the School of Industrial Engineering at the University of Oklahoma. She received her PhD in Industrial Engineering from Arizona State University. Dr. Rhoads teaches engineering statistics and quality engineering courses. Her research interests are diverse in that she studies effective means of learning in the engineering classroom, including the incorporation of web-based learning in statistics. She is interested in not only the cognitive domain of learning, but also the affective domains and how to assess each. In addition, she researches recruitment and retention issues. She can be contacted by e-mail at <teri.rhoads@ou.edu>.