01 Nov 2018

Multi-Objective Evolutionary Neural Network to Predict Graduation Success at the United States Military Academy

Gene Lesinski

Steven Corns
Missouri University of Science and Technology, cornss@mst.edu

Recommended Citation

G. Lesinski and S. Corns, "Multi-Objective Evolutionary Neural Network to Predict Graduation Success at the United States Military Academy," *Procedia Computer Science*, vol. 140, pp. 196-205, Elsevier B.V., Nov 2018.
The definitive version is available at https://doi.org/10.1016/j.procs.2018.10.329

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 140 (2018) 196–205

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Cyber Physical Systems and Deep Learning, CAS 2018, 5 November – 7 November 2018, Chicago, Illinois, USA

Multi-objective Evolutionary Neural Network to Predict Graduation Success at the United States Military Academy

Gene Lesinski[a,*], Steven Corns[b]

[a]*United States Military Academy, West Point, NY 10996, USA*
[b]*Missouri University of Science and Technology, Rolla, MO 10996, USA*

Abstract

This paper presents an evolutionary neural network approach to classify student graduation status based upon selected academic, demographic, and other indicators. A pareto-based, multi-objective evolutionary algorithm utilizing the Strength Pareto Evolutionary Algorithm (SPEA2) fitness evaluation scheme simultaneously evolves connection weights and identifies the neural network topology using network complexity and classification accuracy as objective functions. A combined vector-matrix representation scheme and differential evolution recombination operators are employed. The model is trained, tested, and validated using 5100 student samples with data compiled from admissions records and institutional research databases. The inputs to the evolutionary neural network model are used to classify students as: graduates, late graduates, or non-graduates. Results of the hybrid method show higher mean classification rates (88%) than the current methodology (80%) with a potential savings of $130M. Additionally, the proposed method is more efficient in that a less complex neural network topology is identified by the algorithm.

Keywords: *Evolutionary Algorithms, Neural network; Multi-objective Evolutionary Algorithms; enrollment management; student retention*

1. Introduction

All colleges and universities are concerned with student graduation rates and retention. These data are typically used by organizations like Forbes and US News and World Report as proxy indicators of school quality which indirectly impact the institution's bottom line. Graduation and retention rates are particularly important at the United States Military Academy where a retention loss is ultimately a loss to Army officer end strength. Each year, more

10.1016/j.procs.2018.10.329

than 15,000 candidates, from all 50 states, apply for admission to West Point. Approximately 1,200 applicants are accepted each year and receive the equivalent of a four year full scholarship with a Government Accounting Office (GAO) estimated value of $327,000 [1]. Significant effort is applied to graduate a majority of students within four years to satisfy Army officer manning requirements. Recently there has been a spike in the number of first term course failures for entering freshmen at West Point. This has generated interest in reexamining the decision criteria and models that inform admissions decisions. Given the magnitude of commitment associated with admission and the emphasis on four year completion, it is important to closely examine and periodically revalidate the criteria used to make these important admission decisions. Accurately modeling graduation success can ultimately improve graduation rates, increase student retention, reduce late graduation, and reduce first-term course failures. An accurate prediction model can both inform admission decisions as well as identify students requiring remediation. In this research a pareto-based, multi-objective evolutionary algorithm, utilizing the Strength Pareto Evolutionary Algorithm (SPEA2) [20] fitness evaluation scheme simultaneously evolves connection weights and identifies the neural network topology using network complexity and classification accuracy as objectives. The methodology utilizes nine selected input variables to model and classify student graduation status to inform admission decisions and identify opportunities for required remediation.

1.1. Related Research

In studies of college graduation success, vast amounts of research are focused on identification of significant predictor variables/factors as well as different mathematical models utilizing these factors to predict successful completion of college. There are numerous studies in the literature regarding factors that may predict successful college graduation. These factors are generally divided into pre-admission and post-admissions considerations. Pre-admissions factors can be further categorized as academic and non-academic. Academic pre-admission factors often include, high school rank, high school grade point average, and standardized test scores. Social economic status (SES), parental education, faculty references, and high school extra-curricular involvement are common non-academic factors. Mathematical modeling approaches include regression, Bayesian belief networks, discriminant analysis, support vector machines, and neural networks among numerous others. Most recently, evolutionary algorithms have been applied to similar problems.

1.2. Graduation Prediction Factors

Burton and Ramist found that the best combination of SAT scores to be the best predictor of graduation success [2]. Geisler and Santelices conclude that high school GPA was not only the best predictor of first year grades but also for degree completion [3]. Niu and Tienda argue that another measure of high school achievement, high school rank, is a better predictor of college performance than standardized test scores [4]. Black, et al. found a significant correlation between high school quality and student success at college and believe that high school achievement should be adjusted relative to high school quality [5]. Some examined non-academic indicators of college success include social economic status (SES), parental education, faculty references, and high school extra-curricular involvement. Several sources note the strong correlation between parental level of education and the propensity to attend college. Additionally, Nelson identified a significant relationship between parental education and student college success [6]. Willingham identified faculty references and high school activity involvement as two significant non-academic indicators of college success [7]. In this research, high school rank in conjunction with high school quality, SAT scores, parental education, high school faculty assessments, candidate activity scores, and time since high school are used as factors to predict college graduation success.

1.3. Graduation Prediction Models

Within the literature there are also a wide variety of modeling approaches applied to prediction of college graduation. Bowen and Bok utilized a logistic regression model to predict graduation within six years using gender, ethnic group, SES, selectivity of the college, SAT scores and high school records as predictive factors [8]. Kanarek achieved successful results using discriminant function analyses to classify students into graduates and non-graduates with a combination of pre-admission and post-admission factors [9]. Yingkuachat, et al. use Bayesian belief networks to determine important college graduation success prediction variables with resulting high prediction accuracy [10]. Karimi, et al. utilize a hybrid decision tree and cluster analysis model to identify at-risk college students [11]. Barker, Trafalis, and Rhoads use neural networks and support vector machines to classifying successful student graduation rates at a 4-year institution utilizing student demographic, academic, and attitudinal information [12]. Lesinski. et al. employed a multi-layer feedforward neural network with backpropagation learning to predict graduation success at the United States Military Academy [13].

1.4. Evolutionary Algorithms and Multi-Objective Evolutionary Algorithms

Evolutionary algorithms (EA) are a biologically inspired non-gradient optimization technique that allows the rapid and efficient exploration of vast solution space [14] and have been successfully applied to multiple problem domains [15]. Major components of an EA are a population of potential solutions (chromosomes), mechanisms to select, mate, and exchange portions of solutions with each other, and a means to evaluate solution fitness. The advantages of EA use are that they work well for objective functions that are noisy or not smooth, avoid being trapped in local optimal solutions, can search multiple points in the solution space simultaneously, and can accommodate very large numbers of objective parameters and decision variables. Multi-objective EAs (MOEA) are a special category of EA that consider multiple, often conflicting, objective functions. In this work we attempt to both maximize classification accuracy while minimizing network complexity. Several literature sources summarize, classify, and critique various MOEA [16,17,18]. Approaches to accommodate multiple objectives in EAs can generally be categorized as aggregation-based or pareto- based. Pareto-based MOEA approaches utilize various dominance measures as a means to evaluate solution fitness and guide exploration of the solution space. Pareto optimality excludes from consideration all alternatives or solutions that provide no additional value over other solutions. A primary advantage of the pareto-based approach is the generation of a collection of near pareto optimal, non-dominated solutions which allow decision makers to examine and compare the cost vs. benefits of solutions within the non-dominated solution set. Three current, commonly benchmarked, pareto-based MOEA are Non-Dominated Sorting Genetic Algorithm (NSGAII), Pareto Envelope based Sorting Algorithm (PESA), and Strength Pareto Evolutionary Algorithm (SPEA2) [18,19,20]. The primary differences between these pareto-based approaches are their fitness assignment scheme, diversity mechanism, and use of an external archive. In this work the SPEA2 pareto-based MOEA approach is applied.

1.5. Evolutionary Neural Networks

Evolutionary multi-objective evolutionary neural networks attempt to overcome the difficulty of determining an appropriate network architecture by combining an EA with a neural network. The EA simultaneously develops the neural network topology and “trains” the weights of the network. Brill et al. use a genetic algorithm strictly for input feature selection for a classification neural network. The genetic algorithm fitness function is a linear combination of the classification error and number of features [21]. Maniezzo introduced a parallel genetic algorithm with a novel genetic operator and a granularity encoding scheme to derive the topology and weights for a neural network and applied the methodology to Boolean function learning [22]. Yao presents and analyzes different combinations of EAs and Neural Networks to include EAs that evolve connection weights, input features, learning rules, and architectures with specific discussion of recombination operators and their impact on performance [23]. Fieldsend and Singh apply a pareto-based MOEA using multiple error measures to evolve the weights, topology, inputs, and connectivity within the network in addition to a bootstrap training methodology [24]. Abbass developed a multi-objective evolutionary neural network utilizing competing objectives of minimization error and the number of hidden neurons. Pareto Differential Evolution (PDE), a variant of EA, supplemented by local search via backpropagation was employed to overcome the typical slowness in convergence [25]. Du et al. apply a multi-objective evolutionary neural network to a short-term replenishment forecasting problem. The MOEA evolves the connection weights and the number of hidden nodes within the network. K-fold cross-validation on the training samples is utilized as an error term instead of root mean square error to overcome problems presented by a small data sample [26]. Giustolisi and Simeone employ three competing objectives: number of model inputs, number of hidden neurons, and generalized error against a validation data set to predict ground water levels utilizing monthly rainfall [27]. In this research, a pareto-based, multi-objective evolutionary algorithm based upon the SPEA2 fitness evaluation scheme is used to simultaneously evolve connection weights and identify the network topology using network complexity and classification accuracy as objective functions. The representation scheme and differential evolution recombination operators presented by Abass [25] are employed. This methodology is used to classify student applicants as graduates, non-graduates, or late graduates using 5100 accepted student records with approximately 1300 accepted students per class.

1.6. Current Methodology Employed

West Point currently uses a proprietary linear combination of five factors to quantify candidate quality and inform admissions decisions. The five factors are CEER, Faculty Assessment Score (FAS), Candidate Activity Score (CAS), Candidate Leadership Score (CLS), and Candidate Fitness Assessment (CFA). Each factor score ranges from a possible 200 minimum to 800 maximum. The CEER is a score intended to capture academic performance/potential and includes factors such as: HS rank, HS quality, SAT/ACT scores. FAS is a score assigned based upon 1 x English, 1 x Mathematics, and 1 x Science teacher assessment of academic potential. CAS is a score assigned based upon depth and breadth of extra-curricular activities. CLS is a score assigned based upon demonstrated leadership duties and activities CFA is a score based upon a standardized physical fitness test. These five factors are combined to formulate a Whole Candidate Score (WCS). A general risk level is established for each individual factor (~500) as well as the WCS (~5200). These levels were determined by a series of linear regression equations. If a candidate has a risk in a sub-factor or WCS, additional analysis is conducted by the admissions committee to make a final determination of qualification status or remediation requirements. The current methodology has approximately an 80% classification rate (i.e. non-grads and late grads versus actual graduates). With a cost of approximately $327K per misclassification that equates to over $300M misclassification cost for the classes of 2012-2015.

2. Data Requirements and Pre-Processing

Since academic failure is the primary reason for departure or extended duration stay, academic indicators are primarily considered as model inputs. Academic indicators utilized in this research include: HS rank, HS quality, SAT/ACT Math scores, SAT/ACT English scores, and Faculty Assessment scores. Additionally, we include other factors that previously presented research indicates a high correlation to graduation rates/success: HS extra-curricular activity, parent education, and time since HS. The major outputs of the model are whether a student graduates, does not graduate, or graduates late. Figure 1 highlights the model inputs and outputs.

2.1. Data Pre-Processing

The required data for this research was collected from two primary sources: West Point admissions database and the annual Cooperative Institutional Research Program (CIRP) survey. The CIRP survey data provides candidate parent’s level of education. All other data elements were retrieved from the West Point admissions database. The combined data was “cleaned” by screening for errors and missing data elements. Records with missing data elements or errors were removed with no significant decrease is the overall number of data samples or change in the underlying data set. After cleaning the data set there were 5100 data samples from the Classes of 2012-2015 which consisted of 9 input variables and 3 outputs.

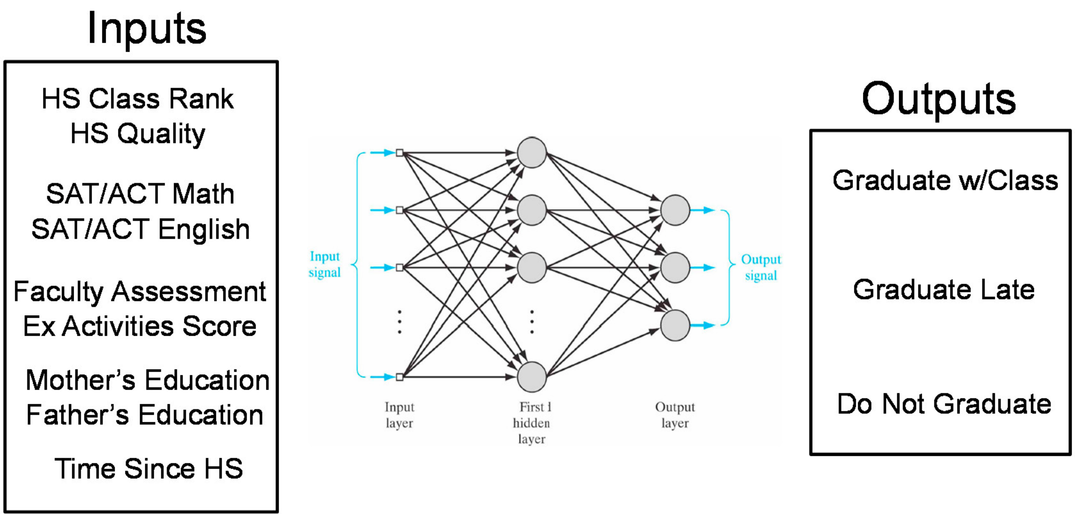


Fig. 1. Model inputs and outputs.

The major data pre-processing tasks required prior to training the neural network for the data previously described were: conversion of SAT/ACT test scores to national percentiles, conversion of categorical variables to binary values, and normalization of numerical data elements. The College Board provides a mapping of SAT/ACT scores to national percentile values [28]. Of the nine input variables, five are categorical (HS Rank, HS Quality, Mother's Education, Father's Education, and Years since HS). To convert the categorical variables into binary representations requires transforming a categorical variable into an equivalent number of binary variables. Binary representation of categorical variables was chosen to facilitate future reduction of model variables while minimizing the impact on model structure. The final data pre-processing step is standardizing the SAT data, Faculty Assessment scores, and Activity scores.

3. Modeling Approach

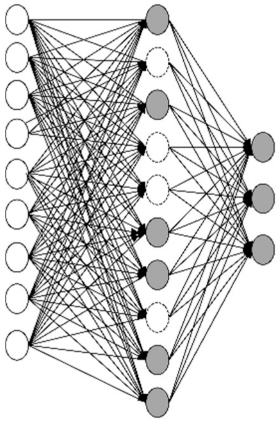
In this research an evolutionary, multi-layer neural network with one hidden layer of neurons is employed. After pre-processing, there are 39 model inputs and 3 model outputs. The number of hidden neurons are varied, using a pareto-based multi-objective evolutionary algorithm (MOEA), up to a maximum of 70. The input to hidden layer weights and the hidden to output layer weights are also trained by the MOEA. Although not the focus of this research, the methodology could also be used to train the output layer and optimize the number of input nodes. The hyperbolic tangent activation function is employed for the hidden layer while the hardlim activation function is used for the outputs. As noted earlier, the MOEA modeling paradigm requires a representation scheme, a fitness evaluation methodology, a selection technique and recombination operators. These essential components are discussed in the sections that follow.

3.1. Representation Scheme

The representation scheme employed for this research combines a binary string component for the number of hidden neurons and 2 matrices for representing the input to hidden and hidden to output layer weights. This representation scheme is highlighted in Figure 2. The solid nodes in the hidden layer are those that are active and therefore are represented as ones in the binary string portion of the chromosome. Note also, the two matrices below. The first is the input to hidden layer weigh matrix for the chromosome and the second is the hidden to output layer weights.

Hidden Layer Neurons

1	0	1	0	0	1	1	0	1	1



Input - Hidden

0.87	0.51	0.08	0.88	0.27	0.24	0.03	0.79	0.64	0.9
0.12	0.57	0.31	0.85	0.15	0.75	0.99	0.85	0.87	0.86
0.35	0.92	0.6	0.47	0.71	0.9	0.65	0.47	0.52	0.91
0.17	0.2	0.7	0.21	0.45	0.09	0.36	0.31	0.76	0.01
0.09	0.18	0.67	0.83	0.55	0.56	0.16	0.41	0.2	0.74
0.99	0.63	0.92	0.71	0.35	0.43	0.6	0.99	0.84	0.65
0.72	0.82	0.7	0.99	0.86	0.71	0.3	0.72	0.66	0.14
0.15	0.75	0.32	0.2	0.38	0.37	0.41	0.24	0.58	0.17
0.31	0.92	0.83	0.3	0.04	0.76	0.91	0.92	0.11	0.8

Hidden - Output

0.2	0.61	0.17
0.66	0.67	0.82
0.33	0.01	0.3
0.84	0.32	0.38
0.35	0.05	0.1
0.5	0.78	0.3
0.38	0.81	0.01
0.16	0.77	0.39
0.35	0.24	0.23
0.97	0.46	0.89

Fig. 2. Representation Scheme.

3.2. Fitness Evaluation

A multi-objective approach is applied in this work. Specifically, two objectives are examined simultaneously; classification accuracy and network complexity. The goal is to minimize neural network complexity while maximizing classification accuracy. The number of hidden layer neurons is used as a measure of complexity. Larger, more complex networks suffer from the curse of dimensionality, require additional computational storage and processing, and typically have inferior generalization capability. Classification accuracy is the percentage of correctly classified accepted students (grad, no grad, late grad). MOEA transform multiple, in this case two, performance measures into a single (non-aggregated) measure of fitness. The most common approach is pareto-based. A vector of decision variables $\vec{x}^* \in F$ is Pareto optimal if there does not exist another $\vec{x} \in F$ such that $f_i(\vec{x})$ such that $f_i(\vec{x}) \leq f_i(\vec{x}^*)$ for all i = 1,….,k and $f_j(\vec{x}) < f_j(\vec{x}^*)$ for at least one j. This work applies the pareto-based fitness scheme introduced by Zitzler and Thiele – Strength Pareto Evolutionary Algorithm 2 (SPEA2) [20]. SPEA2 employs a count and rank-based dominance fitness measure to quantify the quality of candidate solutions. As such, a strength (S) and raw fitness (R) score are calculated for each solution.

$$S(i) = \left|\{\, j \,\middle|\, j \in P_t + A_t \wedge i \succ j \,\}\right| \tag{1}$$

The strength score (S), highlighted in Equation 1, indicates the number of other solutions a particular solution dominates. In equation 1 and 2, P_t is the population at time t and A_t is the archive at time t. A solution, i, dominates another, j, if better accuracy is generated with less or equal network complexity compared to the competing solution.

$$\mathrm{R(i)} = \sum_{\mathrm{j} \in \mathrm{P_t} + A_t,\ \mathrm{j} \succ \mathrm{i}} \mathrm{S(j)} \tag{2}$$

The raw fitness (R) score, highlighted in Equation 2, indicates the total number of other solutions that dominate a particular solution. Solutions with an R score of zero are non-dominated solutions. Solutions are then further differentiated by adding a nearest neighbor density factor to the raw fitness score. SPEA2 utilizes a nearest neighbor density estimate to both fine tune fitness and delete excess non-dominated solutions from the archive. The calculation of the density estimate (D) is shown in Equation 3, where K is the square root of the sum of the population size and the archive size. Note that solutions with a larger distance $\boldsymbol{\sigma_i^k}$ to the kth nearest neighbor will have a smaller density score. The density score (D) is used to fine tune the solution fitness score. Solutions with equal raw fitness scores are differentiated by their density scores (smallest density is preferred).

$$\boldsymbol{D(i)=\frac{1}{\sigma_i^k+2}} \quad \boldsymbol{where} \quad \boldsymbol{k=\sqrt{N_p+N_A}} \tag{3}$$

The final fitness of each solution or chromosome is the sum of the density estimate (D) and the raw fitness (R) as indicated by Equation 4 where $\boldsymbol{N_p}$ is the size of the population and $\boldsymbol{N_A}$ is the size of the archive.

$$\boldsymbol{F(i)=R(i)+D(i)} \tag{4}$$

3.3. Selection Scheme

The selection mechanism within MOEA selects fit members from the population and places them in a mating pool for subsequent pairing and exchange of genetic material. There are several selection schemes highlighted in the literature to include: fitness proportionate based, stochastic uniform selection, rank-based, and tournament-based selection. Tournament-based selection is employed in this work. In tournament-based selection, a number of individuals are selected from the population, the one with the highest fitness is selected and placed in the mating pool. The tournament size can vary from 2 to the size of the population. The larger the tournament size the greater the selection pressure which must be carefully balanced with maintaining diversity in the mating pool and subsequent population. A tournament of size 2 is used in this work.

3.4. Recombination Operators

Evolutionary algorithms employ recombination operators (crossover and mutation) as a means to exchange genetic material between two or more fit population members (parents) to create more fit offspring (children). Crossover exchanges major portions of genetic material between mating parents and occurs at a fairly high probability (0.7 to 1.0) while mutation is intended to reintroduce lost genetic material and help the EA escape from a local min/max. Mutation typically occurs with a small probability (< 0.3). This research employs differential evolution crossover and mutation schemes which are highlighted in Equations 5 -10. A primary parent $\propto_1$and two supporting parents $\propto_2$ and $\propto_3$ are selected to mate to produce one child or offspring. With some probability, the weight matrix of $\propto_1$ is perturbed by adding to it a multiple N(0,1) (i.e. a variable distributed with a standard normal distribution which has a mean of 0 and variance of 1) of the difference between the two supporting parent ($\propto_2$ and $\propto_3$) weight matrices. If the probability is not met, the weight matrix of $\propto_1$ remains the same for the child. Equation 7 highlights a similar scheme for the crossover recombination operator for the binary string ρ.

With some probability (crossover probability)

$$\boldsymbol{w_{ih}^{child} \leftarrow w_{ih}^{\propto_1} + \mathrm{N}(0,1)(w_{ih}^{\propto_2}\text{-}w_{ih}^{\propto_3})} \tag{5}$$

Otherwise

$$w_{ih}^{child} \leftarrow w_{ih}^{\propto_1} \tag{6}$$

$$\boldsymbol{\rho_h^{child}} \leftarrow 1, \text{ if } \boldsymbol{\rho_h^{\propto_1} + N(0,1)(\rho_h^{\propto_2}\text{-}\rho_h^{\propto_3}) \geq .5} \qquad 0, \text{otherwise} \tag{7}$$

As noted earlier, mutation is a small probability event designed to reintroduce lost genetic material and/or to help the EA escape a local min or max. Equations 8-9 highlight the mutation operator for the weight matrices $\propto$. With some probability, the weight matrix of $\propto_1$ is perturbed by adding a standard normal factor. If the probability is not met, the weight matrix of $\propto_1$ remains the same for the child.

With some probability (mutation probability)

$$\boldsymbol{w_{ih}^{child}} \leftarrow \boldsymbol{w_{ih}^{child}} + \mathbf{N(0, mutation\ rate)} \tag{8}$$

Otherwise

$$\boldsymbol{w_{ih}^{child}} \leftarrow \boldsymbol{w_{ih}^{child}} \tag{9}$$

Equation 10 highlights a similar scheme for mutation of the binary string ρ.

$$\boldsymbol{\rho_h^{child}} \leftarrow \mathit{1},\ \mathit{if}\ \boldsymbol{\rho_h^{child}} = \mathbf{0} \qquad\qquad 0\text{, otherwise} \tag{10}$$

3.5. Algorithm Pseudo-Code and Model Summary

The SPEA2 algorithm starts with an initial randomly generated population and an empty external archive. The external archive is an application of EA elitist strategy in which a collection of high quality solutions are maintained and used exclusively for mating of future generations. Typically, the archive size is equal to the population size. However, that is not necessary and in this research we utilize an archive size less than the population size (i.e. 30). Next, the fitness values, which are pareto dominance based measures, are calculated for both the population and archive. The specifics of the fitness measure are detailed in the previous section. All non-dominated solutions from the population and archive are copied to the subsequent archive. If the number of non-dominated solutions exceeds the archive size, excess non-dominated solutions are deleted from the archive based upon on a nearest neighbor density measure presented in the next section. A mating pool is formed using binary tournament selection. Members of the mating pool are randomly paired for recombination of genetic material to form new offspring or solutions. Differential evolution recombination operators are utilized to exchange genetic material between mated pairs of solutions. The above process is repeated until termination criteria, in this case number of generations, are met. At termination, the output of the algorithm is the final archive. Table 1 highlights the key neural network and MOEA parameters used in the model.

4. Results

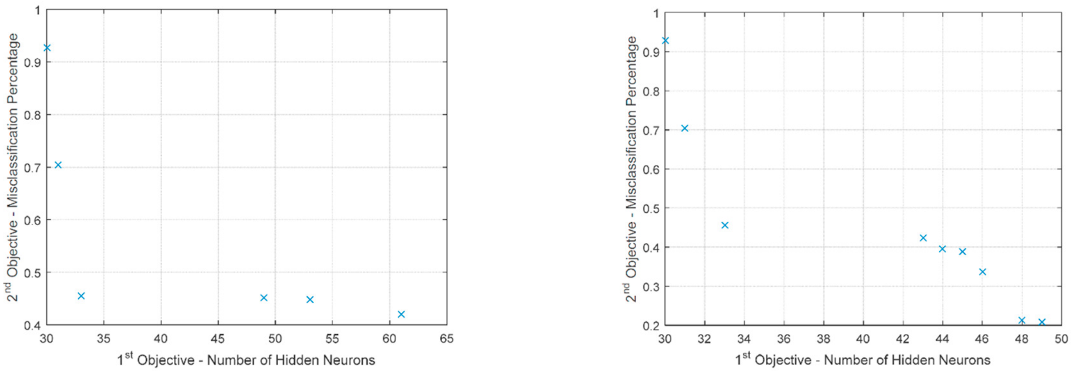
In this section, the performance of a pareto-based, multi-objective evolutionary neural network is examined. This algorithm is used to classify student applicants as graduates, non-graduates, or late graduates using 5100 accepted student records. We compare algorithm performance against a single objective EA and the current admissions prediction accuracy (80%). The performance parameters of interest include classification accuracy and number hidden neurons (complexity). The algorithm is trained on a random sample (3500) of admission records and tested for classification accuracy on a random test set.

Table 1. Key Model Parameters

Neural Network	Input Nodes	43
	Max Hidden Nodes	70
	Output Categories	3
	Activation Function (Hidden Layer)	Tansig
	Activation Function (Hidden Layer)	Hardlim
MOEA	Number of Generations	1000
	Population Size	100
	Archive Size	20
	Probability of Crossover	.4
	Probability of Mutation	.1

4.1. Algorithm Behavior and Results

As a pareto-based MOEA progresses we expect the algorithm to move solutions toward the pareto front and spread them out along the pareto front as the algorithm progresses. Figure 3 highlights this expected behavior on the examined admissions data set. Note that at 10 generations of the algorithm, the highest accuracy (42% misclassification training set) neural network has approximately 62 hidden neurons. After 500 generations, the best network have 20.8% misclassification (training set) with 52 hidden neurons and a competing network with 48 hidden neurons and 21.2% misclassification (training set) accuracy.



3a. 3b.

Figure 3. Algorithm Progression on Training Set (a. 10 Generations b. 500 Generations)

As the upper limit on the number of neuron increases, the design space increases exponentially which increased the algorithm's processing time. To combat this, the upper bounds for the maximum number of hidden neurons were varied. Table 2 highlights the results of the multi-objective evolutionary neural network when the number of neurons are limited (20,30,40,50). The best classification accuracy against the test set achieved was 88% with 37 hidden neurons. These results were achieved when the upper bound on hidden neurons was set to 40.

4.2. Conclusions

This research effort demonstrates application of a pareto-based multi-objective EA to evolve the weights and architecture for a classification neural network. This approach addresses the typical challenge of determining the appropriate neural network architecture while also attempting to train the network. Using 2012-2015 West Point admissions data (5100 total records with 3570 used to train and 1530 used to test), the model was able to achieve 88% classification accuracy with 37 hidden neurons. The high school rank and high school quality were the most important input factors followed closely by parent's level of education. The currently employed West Point methodology has approximately an 80% classification rate (i.e. non-grads and late grads versus actual graduates). With a cost of approximately $327K per misclassification this equates to over $300M misclassification cost for the classes of 2012-2015. Previous work on this same data set utilizing multi-layer feed forward neural network, with manual parameter sweep, required well over 50 hidden neurons and exhaustive search of the neural network architecture space [13] with comparable accuracy results. This work highlights the efficiency gained by using the MOEA to simultaneously evolve connection weights and identify the neural network topology for college admissions applicant classification. There are several potential areas for improvement in the model that may achieve better results. First, the model processing time became burdensome as the upper limit of hidden neurons was increased. A potential upper and lower hidden neuron bound may assist in refining the architecture search space and improve processing time. Second, batching of the training samples may improve processing time and classification accuracy. Memory use and allocation can be improved-a fixed representation scheme was used and is computationally inefficient. Adoption of a variable length representation for both number of hidden neurons and weight matrices could greatly increase processing speed. The proposed technique may also be used to train the output layer and optimize the number of input nodes.

Table 2. Model Results with Varying Upper Limits for Hidden Neurons

Max Hidden Neurons	Class. Accuracy (Train)	Class. Accuracy (Test)	Hidden Neurons
20	79.2%	87%	9
30	76.8%	83%	25
40	79.4%	88%	37
50	79.6%	85%	34

References

1. GAO Report 03-1000, September 2003.
2. Burton, N. W., & Ramist, L., Predicting success in college: SAT studies of classes graduating since 1980, 2001.
3. Geiser, S., & Santelices, M. V. Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. Center for studies in higher education, 2007.
4. Niu, S. X., & Tienda, M. Testing, ranking and college performance: does high school matter. Princeton University,2009.
5. Black, S. E., Lincove, J., Cullinane, J., & Veron, R. Can you leave high school behind?. Economics of Education Review, 46, 52-63, 2015.
6. Nelson, J., Impact of Parent Education on Student Success. Online Submission, 2009.
7. Willingham, W. Success in college: The role of personal qualities and academic ability. New York: College Board, 1985.
8. Bowen, W. and Bok, D., The shape of the river. Princeton, NJ: Princeton University Press, 1998.
9. Kanarek, E., Exploring the murky world of admissions predictions. Paper presented at the Annual Forum of the Association for Institutional Research, Baltimore, MD., 1989.
10. Yingkuachat, J., Praneetpolgrang, P., & Kijsirikul, B. An application of probabilistic model to the prediction of student graduation using bayesian belief networks. Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology Association of Thailand (ECTI Thailand), 63-71, 2007.
11. Karimi, A., Sullivan, E., Hershey, J., Moon, S., A Two-Step Data Mining Approach for Graduation Outcomes, 2013 CAIR Conference, 2013.
12. Barker, K., Trafalis, T., & Rhoads, T., Learning from student data. In Systems and Information Engineering Design Symposium, Proceedings of the 2004 IEEE, pp. 79-86, 2004.
13. Lesinski, G., Corns, S., Dagli, C. Application of an artificial neural network to predict graduation success at the United States Military Academy. Procedia Computer Science, 95, pp. 375-382, 2016.
14. Whitley, D., "A genetic algorithm tutorial," Statistics and computing, vol. 4(2)
15. Murata, T. and Ishibuchi, H., “MOGA: multi-objective genetic algorithms,” Evolutionary Computation, vol. 1, pp. 289-294, November 1995.
16. Coello, C.A., "Evolutionary multi-objective optimization: a historical view of the field," IEEE Comput. Intell. Mag., Vol. 1(1), pp. 28-36, 2006.
17. Konak, A., “Multi-objective optimization using genetic algorithms: A tutorial,” Reliability Engineering & System Safety, vol. 91(9), pp. 992-1007, 2006.
18. Corne, D., “The Pareto Envelope-Based Selection Algorithm for Multi-objective Optimisation,” Proceedings of the 6th International Conference on Parallel Problem Solving from Nature, London, pp. 839-848, 2000.
19. Deb, K., “A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II,” Lecture notes in computer science, 1917, pp. 849-858, 2001.
20. Laumanns, M., “SPEA2: Improving the strength Pareto evolutionary algorithm,” Swiss Federal Institute of Technology, Zurich, Switzerland, Tech. Rep. TIK-Report 103, May 2001.
21. Brill, F. Z., Brown, D. E., & Martin, W. N. (1992). Fast generic selection of features for neural network classifiers. IEEE Transactions on Neural Networks, 3(2), 324-328.
22. Maniezzo, V. (1994). Genetic evolution of the topology and weight distribution of neural networks. IEEE Transactions on neural networks, 5(1), 39-53.
23. Yao, X. (1999). Evolving artificial neural networks. Proceedings of the IEEE, 87(9), 1423-1447.
24. Fieldsend, J. E., & Singh, S. (2005). Pareto evolutionary neural networks. IEEE Transactions on Neural Networks,
25. Abbass, H. A. (2003). Speeding up backpropagation using multiobjective evolutionary algorithms. Neural Computation, 15(11), 2705-2726.
26. Du, W., Leung, S. Y. S., & Kwong, C. K. (2015). A multiobjective optimization-based neural network model for short-term replenishment forecasting in fashion industry. Neurocomputing, 151, 342-353.
27. Giustolisi, O., & Simeone, V. (2006). Optimal design of artificial neural networks by a multi-objective strategy: groundwater level predictions. Hydrological Sciences Journal, 51(3), 502-523.
28. College Board. http://www.collegeboard.com/prod_downloads/highered/ra/act/ACT_SAT_PercentileRanks.pdf