AC 2012-3179: EFFECT OF CLASS ABSENTEEISM ON GRADE PERFORMANCE: A PROBABILISTIC NEURAL NET (PNN)-BASED GA-TRAINED MODEL

Dr. Mohammed E. Haque P.E., Texas A&M University

Mohammed E. Haque, Ph.D., P.E., is a professor of construction science at Texas A&M University at College Station, Texas. He has more than 20 years of professional engineering experience in analysis, design, and investigation of building, bridges, and tunnel structural projects for various city and state governments as well as private sector. Haque is a registered Professional Engineer in the states of New York, Pennsylvania, and Michigan, and a member of ASEE, ASCE, and ACI. Haque received a B.S.C.E. from Bangladesh University of Engineering and Technology, a M.S.C.E. and a Ph.D. in civil/structural engineering from New Jersey Institute of Technology, Newark, N.J. His research interests include fracture mechanics of engineering materials, composite materials and advanced construction materials, architectural/construction visualization and animation, computer applications in structural analysis and design, artificial neural network applications, knowledge based expert system developments, application based software developments, and buildings/infrastructure/bridges/tunnels. inspection and database management systems.

Effect of Class Absenteeism on Grade Performance: A Probabilistic Neural Net (PNN) based GA trained model

Abstract

Most faculty inherently believe that students who frequently miss class significantly increase their likelihood of poor grades by doing so. The purpose of this research was to develop a Probabilistic Neural Net (PNN) based Genetic Algorithm to assess the relationship between absenteeism and student grade performance in a structural systems course taught by the author. The model was trained to classify the outcomes (pass/fail) of 130 students using records of class attendance and end-of-course final grades. The relative importance/weight of attendance on final grades was then determined. It was found that course attendance and grade performance were positively correlated. The model was then used to accurately predict the success rate of a new group of 80 students based on provided attendance records. Overall, this research shows that the developed PNN based GA model can be used to predict the outcome of student performance in the structural systems class based on anticipated class absence patterns.

Introduction

It is widely believed in academia, that consistent class attendance is a key factor in the academic success of students. Several studies have quantitatively confirmed this belief, revealing that absenteeism negatively impacts student academic performance[1-5]. Stanca[6] used a large panel data set from an introductory microeconomics course to correlate the effects of attendance on course performance. Overall, the research results indicated that, even after controlling for unobservable student characteristics (such as ability, effort, and motivation), attendance still exhibited a statistically significant and quantitatively relevant effect on student achievement. Marburger[7] investigated the relationship between students' absenteeism and subsequent exam performance in a basic microeconomics course. Records were maintained regarding the specific class sessions missed by students throughout the semester. Records were also kept of the sessions wherein course materials relating to specific multiple-choice test questions were discussed. A qualitative choice model revealed that students who missed class on a particular date were significantly more likely to respond incorrectly to questions relating to that session's material than were students who had been present. Choudhury[8] used a multiple regression analysis to examine the relationship between absenteeism and final grade performance in a construction materials and methods course. External factors including the differences in academic ability, quiz performance, and student gender were also included in the model to determine if absenteeism maintained its role as a statistically relevant factor in grade performance. The findings generated from the analysis indicated that attendance and academic ability of a student positively correlated with student performance in the course.

For many years, the application of artificial neural networks in pattern recognition and classification has been extensively studied. Various network architectures including multilayer perceptron (MLP) neural networks, radial basis function (RBF) neural networks, self-organizing map (SOM) neural networks, and probabilistic neural networks (PNN) have been proposed. Because of its ease of model training and its sound statistical foundation in Bayesian Estimation Theory, PNNs have become effective tools for solving a wide range of classification problems [9]. Kramer, Mckay, and Belina[10] used an array of three PNNs to classify signals of an electrocardiogram (ECG) in biomedical applications; Romero, Touretzky, and Thibadeau[11] applied PNNs to Chinese Optical Character Recognition (OCR). Haque and Sudhakar[12] applied ANN Back-propagation (BP) to predict fracture toughness in micro alloy steel.

The author believes that an artificial neural network (ANN) model can similarly be trained to classify the correlation between student performance (pass/fail or grades A, B, C, D, F) and external factors. Hence, the author's objective of this research was to implement a Probabilistic Neural Net (PNN) based Genetic Algorithm model to determine the effect of absenteeism on overall student grade performance in his Structural Systems II course.

Research Methodology

Course and Study Population

The author teaches the Structural Systems II (COSC 421) course for senior undergraduate students in the Construction Science Department. The main objective of the course is to teach the analysis, design, and construction perspectives behind structural systems, mainly for reinforced concrete and steel structures. The author uses his course website in addition to textbook and references. Students have access to lecture notes, problem-solving approaches that include logic tables and flowcharts, practice problems, and reference links. Students were required to earn at least a grade "C" to pass this required course. Hence, for this research, the author considered grades "A", "B", and "C" as "PASS" grade, and grades "D" and "F" as "FAIL" grade. The author then assembled the attendance records and final grades of 210 former students over the course of three consecutive semesters (Fall, Spring, and Summer). In each of these semesters, course contents were divided into three segments, and at the end of each segment, students appeared a comprehensive test. Based on the performance of three tests (EXAM-1, -2 and -3) a student received the final grade. Student absences were thus monitored as follows:

Absence Group 1 = Number of absences in segment 1 (first day of semester to day of EXAM-1).
Absence Group 1 = Number of absences in segment 2 (day after EXAM-1 to day of EXAM-2).
Absence Group 3 = Number of absences in segment 3 (day after EXAM-2 to day of EXAM-3).

Probabilistic Neural Net (PNN) based Genetic Algorithm model

Probabilistic neural networks (PNNs) was first proposed by Specht[13] in the early 90's, to fulfil their predominant role as classifiers. By implementing a statistical algorithm called kernel discriminant analysis, PNNs are capable of mapping input patterns to any number of classifications. The basis of the algorithm divides operations into a multilayered feed forward neural network with four layers, (1) Input Layer, (2) Pattern Layer, (3) Summation Layer, and (4) Output Layer. Figure 1 shows a typical PNN architecture. In the model, the input layer distributes data to "neurons" in the pattern layer, and the neuron of the pattern layer computes its output. The neurons of the summation layer then compute the maximum likelihood of a pattern being classified into a particular class by summarizing and averaging the output of all neurons belonging to the same class.

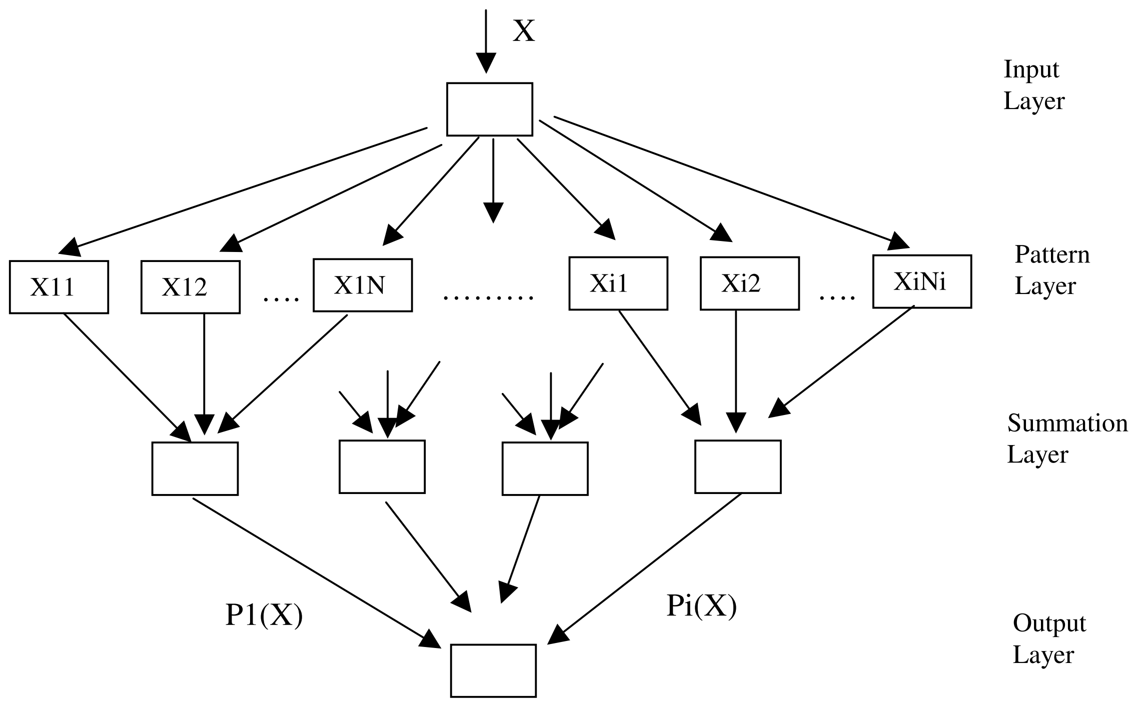

Figure 1. Architecture of a PNN

In this study, the author used "NeuroShell Classifier" (Ward Systems Group, Inc.) to train and evaluate a relevant model. The model was trained using three inputs (three class absence groups) and one output ("PASS" vs."FAIL") using 130 student records. A new set of 80 student records was then used to evaluate the trained model.

Network Training Results

After 152 generation of Genetic algorithm training, the highest number of correct classification plateau was reached, and the model correctly classified 110 records out of 130 (84.62%). Table –1 shows the corresponding network agreement matrix statistics. As shown by the matrix, the network correctly classifies 85.2% (98 of 115) of the "PASS" grades and 80.0% (12 of 15) of the "FAIL" grades. During training it was also observed that the three inputs (absence groups) possessed an unequal bearing on student success. Significance factors for the three inputs were 0.407 for Absence Group 1, 0.315 for Group 2, and 0.278 for Group 3; clearly highlighting the dominant impact Group 1 absences have on student outcome.

Figure 2 shows the Receiver Operating Characteristic (ROC) curves for the network classifications of "PASS" and "FAIL". In the ROC curve the true positive rate (Sensitivity) is plotted as a function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular classification threshold. The area under the ROC curve measures the accuracy of the test. An area of 1.0 represents a perfect test; an area of 0.5 represents a worthless test. The area under the current study's ROC was 0.8351, indicating a suitable test.

Table 1. Agreement Matrix Statistics – Network Training Phase

	Actual "FAIL"	Actual "PASS"	Total
Classified as "FAIL"	12	17	29
Classified as "PASS"	3	98	101
Total	15	115	130
True-pos. ratio	0.8	0.8522	
True-neg. ratio	0.8522	0.8	
Sensitivity	80.0%	85.22%	
Specificity	85.22%	80.0%	

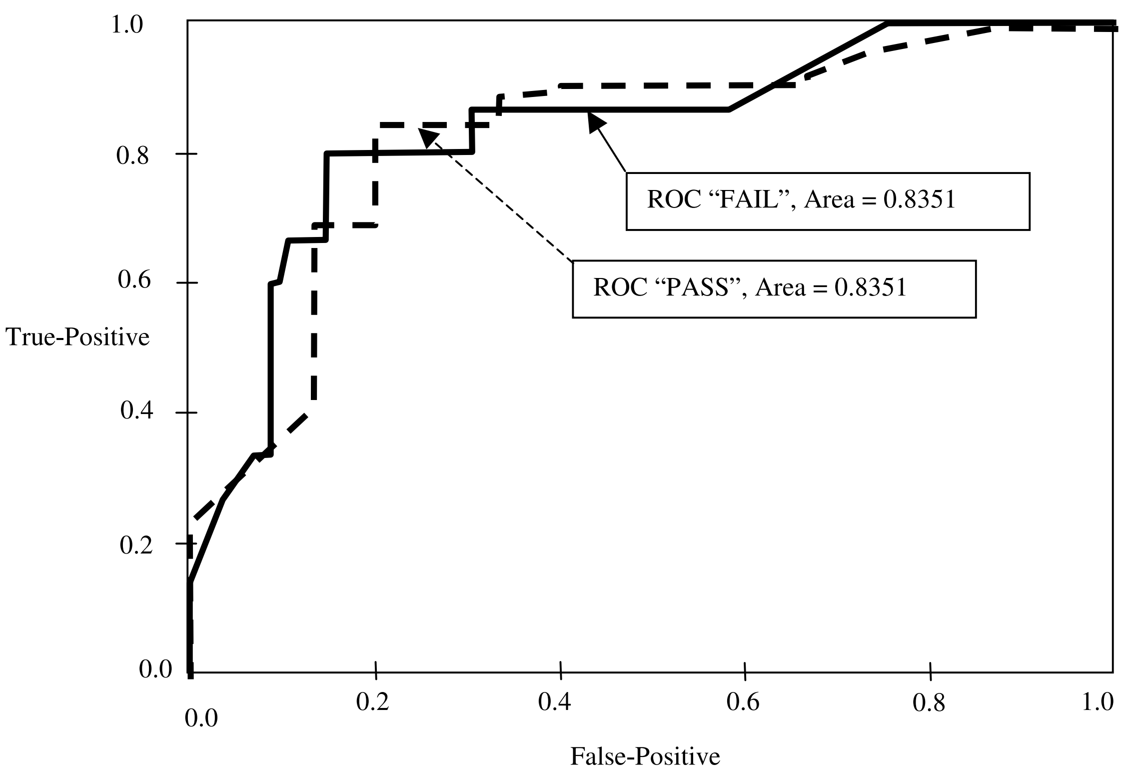

Figure 2. ROC Network Training Curve – True-Positive vs. False-Positive

Network Evaluation Results

The trained network model was then evaluated using a new set of 80 students' attendance records for its ability to predict future grade performances. The trained network model correctly classified 66 records out of 80 (82.50 %). Table –2 shows the network agreement matrix statistics.

Figure 3 shows the Receiver Operating Characteristic (ROC) curves for the network evaluation phase. The area under the ROC of 0.7240 represents a fairly acceptable agreement between the actual and network predicted student performance.

Table 2. Agreement Matrix Statistics – Network Evaluation Phase

	Actual "FAIL"	Actual "PASS"	Total
Classified as "FAIL"	1	10	11
Classified as "PASS"	4	65	69
Total	5	75	80
True-pos. ratio	0.2	0.8667	
True-neg. ratio	0.8667	0.2	
Sensitivity	20.0%	86.67%	
Specificity	86.67%	20.0%	

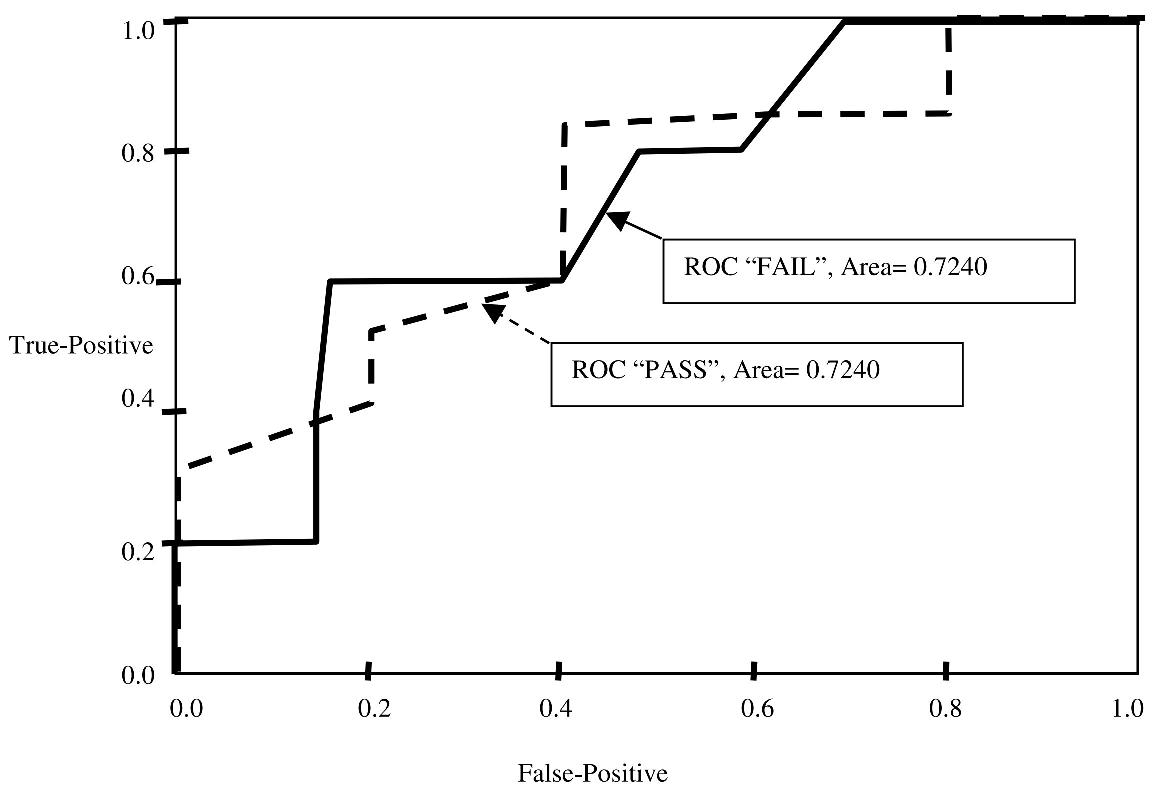

Figure 3. ROC Network Evaluation Curve – True-Positive vs. False-Positive

Trained Network Use and Prediction of Typical PASS/FAIL

Using the NeuroShell Fire Program (Ward Systems Group, Inc), the trained network model can be recalled for future analysis. Figure 4 shows the Input/Output screen. One can input class attendance records and assess the likelihood of future student successes based on anticipated attendance patterns. Table 3 shows a set of typical PASS/FAIL predictions. For example, in Line 1 (Table 3), if a student has a perfect class attendance record, his/her probability of passing the course is 74.7%. Line 5 shows that if a student were absent from the class 50% time in all three segments, his/her chance of passing the course would be only 21.2%. If a student were to miss 25% of class within each course segment as shown in Line 6, the student's margin for passing becomes very narrow (probability of PASS is 54.5%, probability of FAIL is 45.5%).

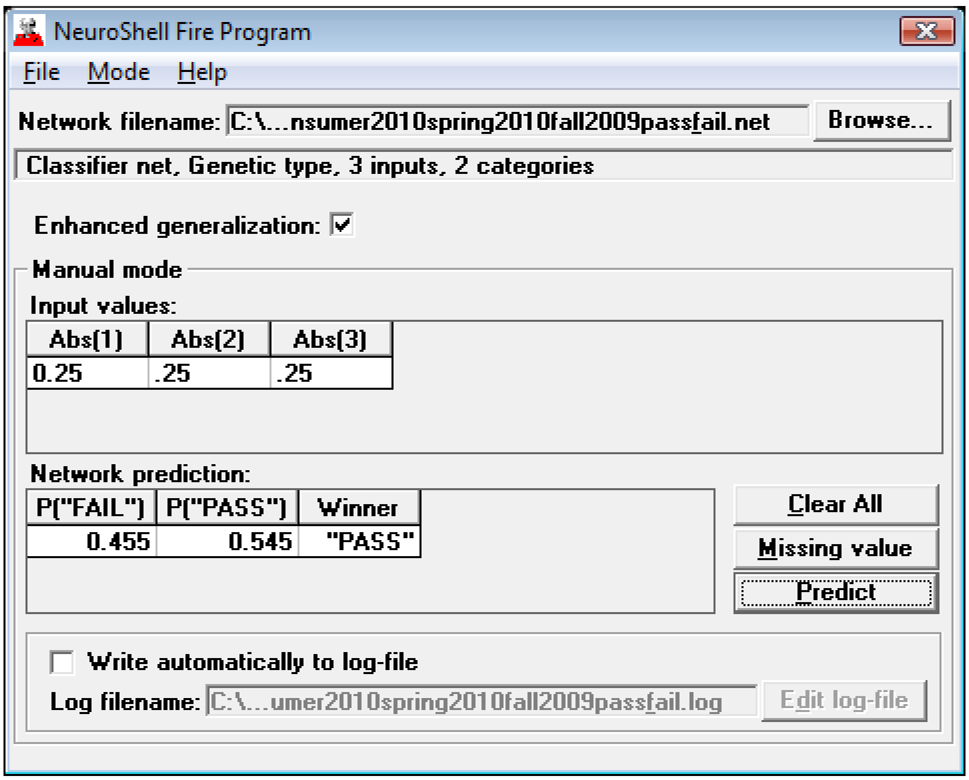

Figure 4. Trained Network Use – Input/output Screen

Table 3. Pass/Fail Prediction

Line	Abs(1)	Abs(2)	Abs(3)	P("FAIL")	P("PASS")	Winner
1	0	0	0	0.253	0.747	PASS
2	1	0	0	0.954	0.046	FAIL
3	0	1	0	0.303	0.697	PASS
4	0	0	1	0.480	0.520	PASS
5	0.50	0.50	0.50	0.788	0.212	FAIL
6	0.25	0.25	0.25	0.455	0.545	PASS

Conclusions

A Probabilistic Neural Net (PNN) based Genetic Algorithm model was trained to classify the pass/fail patterns of 130 students' class attendance records and their grade performance in a structural systems course taught by the author. Absence records were divided into three groups following the course exam schedule. The importance/weight of students' absence sequences on their grade performance was also determined. It was found that attendance in the early phases of the semester had the most significant bearing on final performance. The trained network model was then validated for its predictive capability using a new set of 80 students' attendance records. The area under the Receiver Operating Characteristic (ROC) curves for both network training and evaluation phases indicated satisfactory agreement between actual and network predicted student performance. The study shows the ease by which a trained network can be used to correlate and predict the likelihood of students' academic success on the basis of quantifiable sets of external factors.

Lesson learned and Future Study

The study found that the consistent class attendance is a key factor in the academic success of students, and class absenteeism negatively impacts students' ultimate course results. The results of this study can be used to demonstrate to new students the importance of class attendance, and motivate them not to miss classes in order to be successful in the course. The present study also found that attendance within the first of the three phases of the semester had the most significant bearing on the final performance. In order to make all the phases with nearly equal importance/weight, the author has divided the entire course into four segments with more or less equal amount of course contents, and has been offering since last year.

In the present study, the PNN model was trained to classify the correlation between student performance (PASS/FAIL) and the external factors (three groups of class absence record for entire semester). The grades, "A", "B" and "C" were considered as "PASS", and grades, "D" and "F" as "FAIL". The author believes that the PNN model can be improved to allow predictions at the level of grades - A, B, C, D, and F. To improve the model, the author will be considering in future research additional inputs, such as students past grades on prerequisite courses (Structure I, Mathematics, Physics).

References

1. Park, K.H., & Kerr, P. (1990). Determinants of Academic Performance: A Multinomial Logit Approach. *Journal of Economic Education 21*, 101-111.
2. Romer, D. (1993). Do Students Go to Class? Should They? *Journal of Economic Perspectives 7,* 167-174.
3. Devadoss, S., & Foltz, J. (1996). Evaluation of Factors Influencing Student Class Attendance and Performance. *American Journal of Agricultural Economics 78*, 499-507.
4. Brocato, J., (1989). How much does coming to class matter? Some evidence of class attendance and grade performance. *Educational Research Quarterly, 13(3),* 2-6.
5. Moore, R. (2005). Attendance: Are penalties more effective then rewards? *Journal of Developmental Education, 29(2),* 26-32.

6. Stanca, L. (2006). The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics. *Journal of Economic Education, Volume 37, Number 3 / Summer 2006,* 251 – 266.
7. Marburger, D.R. (2001). Absenteeism and Undergraduate Exam Performance. *Journal of Economic Education, Volume 32, Number 2 / Spring 2001*, 99-109.
8. Choudhury, I. (2008). Predictors of student performance in a construction materials and methods course. *Proc. COBRA: RICS Construction and Building Research Conference*, Dublin, Ireland.
9. Mao, K.Z., Tan, K.C., & Ser, W. (2000) "Probabilistic Neural-Network Structure Determination for Pattern Classification", *IEEE Transactions on Neural Networks, Volume 11*, No. 4, 1009-1016.
10. Kramer, C., Mckay, B., & Belina, J. (1995). Probabilistic neural network array architecture for ECG classification. *Proc. Annu. Int. Conf. IEEE Eng. Medicine Biol., Vol. 17,* 807–808.
11. Romero, R.D., Touretzky, D.S., & Thibadeau, G.H. (1997). Optical Chinese character recognition using probabilistic neural networks. Pattern Recognit., Vol. 3, No. 8, 1279–1292.
12. Haque, M.E., & Sudhakar, K.V. (2002). ANN Back-propagation Prediction Model for Fracture Toughness in Microalloy Steel. *International Journal of Fatigue, Vol. 24, No. 9,* 1003-1010.
13. Specht, D.F. (1990). Probabilistic neural networks. Neural Networks, vol. 3, no. 1, 109–118.