

NEW RESULTS IN COMPUTER AIDED DIAGNOSIS (CAD) OF BREAST CANCER USING A RECENTLY DEVELOPED SVM/GRNN ORACLE HYBRID

WALKER H. LAND, JR.,

Research Professor
Dept. of Computer Science
Binghamton University
Binghamton NY

TIMOTHY MASTERS

President ,TMAIC
Brackney , Pa.

FRANCES ANDERSON

Lourdes Hospital and the Regional
Cancer center
Binghamton, NY

DOUGLAS M. RITZ

Graduate Student
Dept. of Computer Science
Binghamton University
Binghamton, NY

LUT WONG

Graduate Student
Dept. of Computer Science
Binghamton University
Binghamton, NY

DAN MCKEE

Graduate Student
Dept. of Computer Science
Binghamton University
Binghamton, NY

JOSEPH Y. LO

Duke University Medical Center
Duke University
Durham, NC

ABSTRACT

This research consisted of evaluating diagnostic performance results using SVM outputs previously obtained from an integrated Duke/DDSMS USF data set and the GRNN oracle. The SVM kernels used in this research included Additive, Multiplicative, S2000, and Spline kernels. GRNN results are presented for the following combinations of gate variables: age, mass margin (MM), age and MM, and all 6 BIRADS™ indicators plus age. For all experiments, Differential Evolution (DE) was used to train the GRNN. A summary of the DE process is described, independent of the software application. The experiments described in this paper show that the GRNN oracle, with all of the gate variable combinations, performed better than any of the individual SVM kernels alone at or below 98% sensitivity.

INTRODUCTION

Carcinoma of the breast is second only to lung cancer as a tumor-related cause of death in women. For 2003, it has been reported that 211,300 new cases and 39,800 deaths will occur just in the US . It has been proposed, however, that mortality from breast cancer could be decreased by up to 25% provided that all women in appropriate age groups were regularly screened . Currently, the method of choice for early detection of breast cancer is mammography, due to **its general widespread availability, low cost**, speed, and non-invasiveness. At the same time, while mammography is sensitive to the

detection of breast cancer, it has a low positive predictive value (PPV), resulting in costly and invasive biopsies that are only 15%-34% likely to reveal malignancy at histological examination [15].

Computational intelligence has been applied to the problem of breast cancer screening by several researchers. Earlier work by Floyd et al. [5] and Lo et al. [16] demonstrated the applicability of back propagation-trained Multiple Layer Feed forward Neural Networks (MLFNs) to this task. EP-based feed forward networks were applied by Fogel et al. [7-9] and Land et al. [13]. These applications were able to achieve similar or better results than those trained by back propagation, but with much simpler architectures (i.e., fewer nodes in the hidden layer). Evolutionary-Programming (EP) based linear discriminate analysis [10], and an adaptive boosting/EP hybrid [14] also were investigated as potential classification mechanisms. This paper extends the knowledge gained from previous breast cancer research through a description of the development and preliminary evaluation of a new SVM/GRNN oracle hybrid.

THE GENERALIZED NEURAL NETWORK (GRNN) ORACLE

Designing the GRNN Oracle

We have two or more different models, each of which predicts the same scalar output variable [17]. (Extension to multiple outputs is easy, but in practice we never want to do this.) If there are multiple outputs, use a separate oracle for each.) We do not care about the nature of these models, nor do we care about their inputs. In fact, the model inputs do not enter into this discussion in any way. We also have one or more *gate variables* whose values presumably have an effect on deciding which of the competing models is most valid. For our purposes now, we do not care about the individual prediction models. They are assumed to be black boxes that work reasonably well. Our current goal is to design an oracle that used the gate variables to intelligently combine the outputs of the competing models.

EXPERIMENTAL DESIGN TO EVALUATE INTEGRATED HARDWARE/SOFTWARE SYSTEM AND SENSITIVITIES ANALYSIS

This section describes the experiential design approach that will be used. The measures of performance which will be used are: the overall ROC Az index, the specificity and PPV at 95% (5% false positive errors (FPE)), 98% (2% FRE) and 100% sensitivities (0% FPE). As previously described, these MOPs provide a quantities measure of how well the system decreases false negative errors (FNE), while maintaining the 95%, 98%, and 100% sensitivities. K-fold cross will be employed.

Using the GRNN oracle.

We will accomplish data fusion of several learning machine models using the GRNN oracle. This oracle intelligently combines the outputs of several of these

models by using gate variables. The inputs are from several Support Vector Machine models (to be described subsequently). These several SVM learning machines are comprised of the following kernel types: s2000, Gaussian Radial Basis Function (GRBF), polynomial kernels, spline kernels, sum and product kernels whose components are the GRBF and polynomial kernels. Secondly, the gate variables used in this research are (also subsequently described): (1.) age, (2) mass margin (MM) and (3) age and MM. Suppose we have two or more prediction models of any type, each of which predicts the same scalar output variable. Extension to multiple outputs is easy, but in practice we never want to do this. If there are multiple outputs, use a separate oracle for each.)

Our current goal is to design an oracle that uses the gate variables to intelligently combine the outputs of the competing models. Once the expected error of each prediction model is estimated, we use these expected errors to compute the weights for each model. When an unknown case is processed, the gate variables are used by the GRNN to decide which models are likely to be best for this particular case. These models are weighted more heavily than the likely inferior models.

SUPPORT VECTOR MACHINES AND DIFFERENTIAL EVOLUTION

Several references are available that provide extensive information about development of the mathematical foundation of SVMs [3,4,11,12]. The key concepts of SVMs generally are covered in these references and will not be discussed here.

DIFFERENTIAL EVOLUTION

Price and Storn [18] reported on a variation of genetic optimization called *differential evolution*. This variation appears to be much more appropriate than traditional genetic methods when optimizing a multivariate function. It is especially valuable when the scalings in the different dimensions are not commensurate, a situation commonly found in GRNN training using poorly prescaled or highly correlated variables. Unfortunately, differential evolution shares the principal weakness of all stochastic methods in that it can arrive frustratingly close to the global minimum, then fail to converge to the minimum in a reasonable period of time. The problem is that these algorithms generally operate in total ignorance of the local properties of the function being minimized. Sometimes this is out of necessity because the derivatives cannot be computed. More importantly, it is because we must be careful to avoid excessive use of local information if we are to preserve the global quality of the search for the minimum. But when we do have easy access to local information, it often makes sense to make modest use of it. Such is the case when using differential evolution to train a GRNN.

Differential evolution is similar to ordinary genetic optimization in that it starts with a collection of parameter sets that we will call the *source*

population. The individuals comprising this population are combined with each other via *crossover* and subjected to *mutation* to produce the members of the *destination population*. The members of the destination population, taken as a group, are generally expected to be superior to the members of the source population. By repeating this process enough times, the best member of the final population is hopefully close to the global optimum.

There are several important differences between traditional genetic optimization and differential evolution. Probably the most important difference is in the nature of the mutation. In traditional genetic optimization, mutation takes the form of a random perturbation of a fixed type, such as flipping bits in a binary representation of a parameter set, or adding random numbers to individual parameters. The problem with this approach is that it fails to account for the fact that what might be a small perturbation for one parameter might be gigantic for another. Also, random bit flipping can be extremely destructive. Differential evolution avoids these problems by using the source population itself to determine the nature and degree of mutation. It does this by randomly selecting a pair of individuals and computing the difference between their parameter vectors. This difference vector is multiplied by a fixed constant (typically around 0.5 or so) and added to the individual being mutated. When the optimization begins, the average difference will be about the same for all variables being optimized. But as generations pass, the difference will tend to adapt to the natural scaling of the problem. Variables having a large natural scale will be distributed over a larger range in the population, so mutations for these variables will also be relatively large. As convergence approaches, those variables having a narrow and well defined range around the minimum will have small variation among the population members, resulting in their mutations being relatively small. This automatic adaptation significantly improves behavior of the algorithm as convergence nears.

Another important difference is that differential evolution does not involve selection of parents based on fitness. Instead, fitness determines which children are kept. In particular, one parent, called the *primary parent* is selected deterministically: each individual in the source population is chosen as a primary parent exactly once. The other parent, called the *secondary parent*, is randomly chosen. Two other individuals which make up the *differential pair* are also selected randomly. These two are subtracted and their difference is multiplied by a small fixed constant. This scaled difference vector is added to the secondary parent to induce mutation. Ordinary crossover is applied to the primary parent and the mutated secondary parent. The resulting child's fitness is compared to that of the primary parent. The winner becomes a new member of the destination population. This entire process is illustrated in Figure 2.

PRELIMINARY RESULTS

This section describes the experiential design approach that will be used. The measures of performance which will be used are: the overall ROC Az index, the specificity and PPV at 95% (5% false positive errors (FPE)), 98% (2% FRE) and 100% sensitivities (0% FPE). These MOPs provide a quantities measure of how well the system decreases false negative errors (FNE), while maintaining the 95%, 98%, and 100% sensitivities. K-fold cross will be employed.

Several research experiments were performed to evaluate the performance of the separate support vector machine models, and the use of differential evolution to train the GRNN oracle. These families of SVMs were evolved over 40 generations, with a population size of 30, using Evolutionary Programming.

	GRNN	SVM	Difference (GRNN – SVM)	% Improvement
Az	81.62%	81.21%	0.41%	0.51%
Az-90	32.01%	29.61%	2.4%	8.11%
100% Spec	5.56%	3.27%	2.29%	70.03%
100% PPV	48.23%	47.59%	0.64%	1.35%
98% Spec	22.15%	17.85%	4.3%	24.09%
98% PPV	52.67%	51.33%	1.34%	2.61%
97% Spec	24.23%	23.46%	0.77%	3.28%
97% PPV	53.16%	52.81%	0.35%	0.66%
95% Spec	32.38%	29.44%	2.94%	9.99%
95% PPV	56.96%	54.27%	2.69%	4.96%

Table 1: Performance Comparison between SVMs and GRNN oracle

CONCLUSIONS

The experiments just described demonstrated the following: (1.)No measurable performance accuracies difference resulted for variation in the differential evolution crossover constant, (2.)The Az and partial Az are essentially constant for all values of the differential evolution crossover constant, holding at values of 81.6% and 32%, respectively. A partial Az of 32% means that, on the average, approximately 432 women would avoid biopsy when compared to all cases being biopsied, (3.)Produces a PPV of from 48% to 55% as the sensitivity is decreased from 100 to 95%, respectively, (4.) specificity is increased from about 5% to about 36% as sensitivity is decreased from 100 to 95%, and (5.)Improvements from about 0.5% to 24% were observed in the GRNN oracle performance when compared to the SVM diagnostic accuracy. This is not surprising when one observes that SVMs can be trained to a global minimum, given that the learning machine parameters are properly computed.

ACKNOWLEDGEMENTS. This work was supported in part by the New York state breast cancer research and education fund through department of health contract C017939.

REFERENCES

1. J. A. Baker, P. J. Kornguth, J. Y. Lo, M. E. Williford, and C. E. Floyd, Jr., "Breast Cancer: Prediction with Artificial Neural Network Based on BI-RADS Standardized Lexicon", *Radiology*, Vol. 196, pp. 817-22, 1995.
2. *BIRADS, Breast Imaging - Reporting and Data System (BI-RADS)*, American College of Radiology, 1993.
3. C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp. 121-167, 1998.
4. N. Cristianini, and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*, Cambridge University, 2000.
5. C. E. Floyd, J. Y. Lo, A. J. Yun, D. C. Sullivan, and P. J. Kornguth, "Prediction of Breast Cancer Malignancy Using an Artificial Neural Network", *Cancer*, Vol. 74, pp. 2944-98, 1994.
6. D. B. Fogel, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, IEEE Press, Piscataway, NJ, 1995.
7. D. B. Fogel, E. C. Wasson, and E. M. Boughton, "Evolving Neural Networks for Detecting Breast Cancer", *Cancer Letters*, Vol. 96, pp. 49-53, 1995.
8. D. B. Fogel, E. C. Wasson, E. M. Boughton, and V. W. Porto, "A Step Toward Computer-Assisted Mammography using Evolutionary Programming and Neural Networks", *Cancer Letters*, Vol. 119, pp. 93-97, 1997.
9. D. B. Fogel, E. C. Wasson, E. M. Boughton, and V. W. Porto, "Evolving Artificial Neural Networks for Screening Features from Mammograms", *Artificial Intelligence in Medicine*, Vol. 14, pp. 317-326, 1998.
10. D. B. Fogel, E. C. Wasson, E. M. Boughton, V. W. Porto, and P. J. Angeline, "Linear and Neural Models for Classifying Breast Masses", *IEEE Trans. Medical Imaging*, Vol. 17, No. 3, pp. 485-488, 1998.
11. S. Gunn, *Support Vector Machines for Classification and Regression*, ISIS Technical Report, Image Speech and Intelligent Systems Group, University of Southampton, May 14, 1998.
12. S. Haykin, *Neural Networks*, Prentice Hall, New Jersey, 1999.
13. W. H. Land, T. Masters, J. Y. Lo, and D. McKee, "Using Evolutionary Computation to Develop Neural Network Breast Cancer Benign/Malignant Classification Models", *4th World Conference on Systemics, Cybernetics and Informatics*, Vol. 10, pp. 343-7, 2000.
14. W. H. Land, Jr., T. Masters, and J. Y. Lo, "Application of a New Evolutionary Programming/ Adaptive Boosting Hybrid to Breast Cancer Diagnosis", *IEEE Congress on Evolutionary Computation Proceedings*, 2000.
15. J. Y. Lo, J. A. Baker, P. J. Kornguth, J. D. Iglehart, and C. E. Floyd, "Predicting Breast Cancer Invasion with Artificial Neural Networks on the Basis of Mammographic Features", *Radiology*, Vol. 203, pp. 159-163, 1997.
16. J. Y. Lo, J. A. Baker, P. J. Kornguth, and C. E. Floyd, Jr., "Effect of Patient History Data on the Prediction of Breast Cancer from Mammographic Findings with Artificial Neural Networks", *Acad. Radiology*, Vol. 6, pp. 10-15, 1999.
17. T. Masters, and W. H. Land, Jr., "A New Training Method for the General Regression Neural Network", *IEEE International SMC Conference Proceedings*, pp. 1990-5, 1997.
18. K. Price, and R. Storn, "Differential Evolution", *Dr. Dobb's Journal*, pp. 18-24, April, 1997.