CHAPTER 7

CONCLUSION

This dissertation presents a new machine learning model for generating ensembles of classifiers. The primary goal of the research is to improve generalization of constructive networks through network combination. The motivation for this research was inspired by Sollich & Krogh, who suggested the plausibility that ensembles of complex models would exhibit positive results [7]. Other issues addressed throughout this research include the dilemmas associated with architectural decisions, network creation and selection, overfitting, task subdivision, and methods for combining networks. All of these issues are dealt with in the context of their effect on generalization and model robustness.

Existing ensemble methods and classifiers relevant to this study are discussed, along with performance issues and limitations. The relationships of network topology and of modifiable parameters to ensemble construction are investigated. A new membership combination method is proposed for combining the classification or prediction of individual students within an ensemble. The proposed combination method was inspired by Jacob’s call for investigation into combining methods [4] and is
A natural extension of the work by Xu, Krzyzak, and Suen on Averaged Bayes Classifiers [109].

A new neural network ensemble simulation tool is developed for conducting and recording ensemble experiments. The Neural Network Ensemble Simulator provides a facility for generating ensemble students, combining their predictions, and storing results for statistical analysis. Features are available throughout the simulation process to allow flexibility in the ensemble creation, selection, and combination processes. The tool currently provides for an unlimited number of ensemble members. A choice is provided between two learning algorithms and two combination methods. A limited set of modifiable parameters is available. A vast array of statistical data is maintained for all ensemble simulations.

7.1 Specific Findings

The results in Chapter 6 confirm four of the six hypotheses. Cascor Ensembles are shown to outperform ensembles of linear classifiers by 24.47%. The improvement of Cascor Ensembles over Adaline ensembles reflects the greater capacity of nonlinear models over linear models in capturing the complexities of nonlinear data.

Cascor Ensembles are able to outperform their independent Cascor counterparts by 12.93%, at 48.09% greater stability. The combined strengths of the separate networks settling in different local minimum capture unique regions of the feature space. The improvement achieved by combining the Cascor networks supports the idiom that the whole is greater than the sum of its parts.
Ensemble ambiguity is known to be a positive factor in ensemble success [8]. How this ambiguity is achieved is dependent upon network creation and selection. This research presents ambiguity embodied in diverse network topologies. Topological diversity implies student members of varying architectures. Support for interest in architectural diversity among members was inspired by Drucker, Schapire, and Simard [6]. Analysis of the ensemble composition among the results of this study reflect topological diversity in all cases. The results show that the topologically diverse students predict differently on the test data, as they model unique local minima in the feature space. This disagreement reflects network ambiguity. The network ambiguity is confirmed to be a contributing factor to the success of the ensemble models.

Topological diversity is easily achieved by the constructive architecture, yielding student members with varying numbers of hidden units and connections. In 98.78% of the simulations, this diversity resulted in improved performance.

Diversity of networks is typically achieved through task subdivision, i.e. dividing the data in a useful way to achieve unique solutions within the problem space. This is especially critical for linear models. However, given a nonlinear model there are other options for attaining unique solutions. One such method proposed was to achieve network diversity by varying the modifiable training parameters. The benefit of this approach is that no extra work is necessary to achieve a division of the training data. All available training data may be used for learning by each network. Where the network settles will be reliant on where it began and on its parameters for learning, rather than on which data are omitted from training. The given approach employs random varying of
parameter set values. When testing on all available training data, varying modifiable
parameters produces ensembles that outperform individual networks by 15.06%.

The results presented in Chapter 6 are insufficient to confirm Hypothesis 2. Hypothesis 2 states that Cascor Ensembles exhibit improved time-complexity over
backprop ensembles, without compromising overall performance. Further study is
needed to determine the actual timesavings of Cascor Ensembles vs. ensembles of
backprop networks. Future studies will compare the connection crossings required
during training by both methods, for all networks required in generating a successful ensemble.

The results presented in Chapter 6 are also insufficient to confirm Hypothesis 6. Hypothesis 6 states that ensembles for binary classification can benefit from a
combination method that exploits information acquired during the validation process.
The weighted contribution combination method determines the contribution of each
ensemble member by its performance on the validation set. Essentially, each network’s
contribution is inversely proportional to its validation set error. This weighting scheme
offers a soft-decision rule alternative to classification by voting, as well as an alternative
to weighted averaging of prediction. The method does not require any a priori
information about the class distribution of the test set. Additionally, each network's
unique contribution value maintains useful information acquired during validation. The
results presented in this study clearly indicate the validity of the weighted contribution
combination method. Ensembles combined with weighted contribution demonstrate a
2.82% performance increase over ensembles combined by voting. However, the
improvement within each ensemble model, when compared to its own independent networks, is less convincing. More testing is needed to confirm the benefits of the proposed weighted contribution combination method.

7.2 Significance

Ensembles of constructive networks are shown to provide greater generalization than ensembles of linear networks and individual constructive networks when presented with the same data and under identical experimental conditions. The benefits of topological diversity achieved by constructive networks when combined into an ensemble are demonstrated. Ensembles of constructive networks train rapidly and rely on no heuristics for architectural selection. Ensembles of constructive networks generated by subtask or random parameter selection relieve the experimenter from heuristics associated with training parameter value selection as well.

All simulations performed for this research are conducted with the developer’s NNES research tool. NNES provides an ensemble experiment management facility. NNES demonstrates favorable results when applied to the diabetes domain. It should be interesting to observe NNES applied to a variety of different domains. In its current and future versions, NNES will be used extensively for prediction and classification of a variety of real-world problems. Anticipated data repositories include PROBEN1, DELVE, and UCI. Domains of interest for applying NNES and related techniques span such fields as medicine, marketing, geophysics, sociology, economics, and education.
7.3 Limitations

The presented machine learning model is not limited to classification or prediction problems, nor is it limited in size, scope, or application domain. However, the current study is limited to binary classification for simplicity of understanding the model and its behavior. NNES was initially developed to support the current study, with future extensibility in mind. Therefore, the current version of NNES supports only binary classification problems. However, NNES is currently undergoing reconstruction to support multi-classification problems and prediction of continuous-valued outputs.

7.4 Future Research

Future work is planned to explore the potential of the weighted contribution combination method. More testing is planned on binary classification problems, as well as to test the Surrogate Bayes Combination Method on multi-class problems from the PROBEN1 data repository. Future testing may also encompass continuous-valued prediction problems from PROBEN1.

Additions to the NNES machine learning tool are planned. One of the additions will be the inclusion of a variation on backpropagation. The addition of backpropagation will support the local generation of ensembles of backprop networks and recording of experiment statistics and results.

A two-level cross-validation option will also be implemented in NNES. Level one cross-validation (cv1) will perform cross-validation of the training data for network
creation. Level two cross-validation (cv2) will perform cross-validation of the entire data set, providing a performance measure for model creation.

Further enhancements to NNES will provide for greater flexibility at run-time. These will include features for independent training, validation, testing, and prediction. Also to be included are more options for modifying CNNS and NNES training parameters, scripts for running NNES, and a GUI interface (see Appendix A).

7.5 Final Remarks

Ensembles of constructive networks provide a viable alternative to other types of classification systems, namely, linear ensembles and backpropagation ensembles. Cascor Ensembles demonstrate good generalization and stability, without being computationally prohibitive. Constructive ensembles can resolve many of the dilemmas facing researchers when constructing ensembles of classifiers. Such dilemmas would typically include architectural selection, parameter value selection, and task subdivision. A novel combination method is introduced that provides researchers with a new soft-decision rule option for combining classifiers or predictors, regardless of the ensemble architecture or domain features.

Much is learned about the combination of classifiers throughout this investigation. Hypotheses about constructive classification systems are formulated and empirically confirmed. Useful information is derived from the simulations conducted under this research. However, the scope of this study represents but a small picture of the complexities associated with ensemble construction and combination. Future
research is warranted to study the ensemble selection and combination processes, as well as the impact these processes have on time-complexity of model creation and implementation.
APPENDICES

APPENDIX A  –  NNES Documentation ........................................................................... 108
APPENDIX B  –  Interface Modules.................................................................................. 116
APPENDIX C  –  Preliminary Results Based on linear and XOR problems...... 119
APPENDIX D  –  diabetes Data & Adaline Ensemble Results ................................. 132
APPENDIX E  –  Cascor Ensemble Results on diabetes................................................. 145
APPENDIX F  –  Complete Summary of Results on diabetes.................................. 170
APPENDIX G  –  NNES C++ Code .................................................................................. 177