Learning with Linear Hypothesis Classes and Stochastic Global Optimization with Applications to Artificial Neural Networks

Sridev Ramaswamy[†]

Indian Institute of Management Calcutta

2002

Thesis Advisory Committee:

Prof. Asim K Pal (thesis supervisor), Indian Institute of Management Calcutta

Prof. Ambuj Mahanti, Indian Institute of Management Calcutta

Prof. Sushil Khanna, Indian Institute of Management Calcutta

Thesis Examination Committee:

Prof. Asim K Pal, Indian Institute of Management Calcutta

Prof. Rahul Mukherjee, Indian Institute of Management Calcutta

Prof. Swapan K Parui, Indian Statistical Institute Calcutta

Prof. Kalyanmoy Deb, Indian Institute of Technology Kanpur

[†] sridevr@yahoo.com, sridev.ramaswamy@db.com

Thesis Summary

The thesis deals with the following problems in an integrated manner:

a) computational learning problem of model selection with the objective of achieving maximum generalization capability using compound linear hypothesis classes for two-class classification

b) stochastic global optimization for unconstrained functions with an emphasis to high dimensions

c) understanding the role of optimization in the computational learning problem of model selection as well as parameter selection with applications to ANN architecture construction

algorithm. A combinatorial optimization algorithm for threshold neural networks has been suggested. Based on the duality of input space and the parameter space, a zone-crossing algorithm for linear separation has been developed. Two important problems in finance, namely bankruptcy prediction and takeover prediction have been attempted to apply the above concepts to ANN learning (both architecture construction and weight optimization). An extension to a non-linear hypothesis class of hyperspheres has been introduced. The geometric method used in the computation of VC bounds needs to be further strengthened for obtaining exact VC dimensions.

In the area of stochastic global optimization (SGO) for minimizing unconstrained functions, the main contributions of the current work include: development of a simplex based direct search hybrid algorithm for SGO, development of a unified framework to abstract various components of SGO algorithms and development of a family of elite solutions based algorithms to solve global optimization problems.

First, a simplex-based direct search algorithm (NMDT), which is a hybrid of the Nelder-Mead (NM) algorithm and the multi-directional search algorithm of Dennis and Torczon (DT) is proposed. The property of DT to find an improvement in a finite number of steps is used to remedy the output, when NM converges pre-maturely. The algorithm demonstrates its potential in high dimensional problems particularly when the function is noisy or has flat areas. These algorithms are compared theoretically and empirically. Experimental results for a number of benchmark functions upto 500 dimensions show the superior performance of NMDT compared to NM, DT and NMLS (Nelder-Mead with line search). As an illustration to ANN learning, NM has been successfully applied to solve the currency exchange rate prediction problem. NMDT which is likely to perform better than NM has not been tried however. The algorithm NMDT_G is designed to solve the global optimization problem. In NMDT_G, the simplex is changed by perturbing some of its vertices after the DT phase and before the NM phase of the next iteration. NMDT_G is shown to strongly converge to the global minimum using the convergence results of a unified framework for SGO algorithms. Two variants of NMDT_G based on random simplex vertices and annealed simplex vertices are given. The performance of the global variants on the

benchmark problems illustrates the strength of these SGO algorithms to solve a wide range of problems.

In the current thesis, a unified framework for SGO algorithms is proposed. It draws from the basic random search algorithm of Solis & Wets [1981] and generalizes the approach to SGO under one common framework. Important issues that need to be addressed in order to incorporate desired features in a SGO algorithm are abstracted into three components namely **sampling**, which generates a diverse collection of sample points using a suitable probability measure**,** **processing,** which processes the sample points to obtain a candidate solution for a given iteration, and **adaptation**, which determines the outcome of a given iteration based on the candidate point, updates the algorithm parameters for subsequent iterations and maintains a solution history. The framework offers immense flexibility in defining a SGO algorithm through the specification of the different components along with their parameters dynamically. The conditions for the strong convergence of algorithms falling under the framework are obtained. A wide range of SGO algorithms are analyzed along with their convergence characteristics. This analysis has given rise to many possibilities for developing new algorithms including hybrid ones. A few of such algorithms have been suggested here.

Finally, an elite solutions based framework (ESBF) for SGO is presented. In each iteration of an algorithm under ESBF, a set of points is randomly chosen first. From this initial solution set, a subset of elite solutions is chosen based on certain criteria. The elite solutions then become the basis for further search by the algorithm. This framework is a special case of the unified framework, in the sense that the *processing* component involves the choice of elite solutions. Several ESB algorithms are proposed and their convergence characteristics are analyzed. A basic variant of ESB is compared analytically with the multi-start algorithm with respect to the solution quality as well as the number of function evaluations. This analysis indicates how algorithm parameters can be dynamically tuned to facilitate adaptive search. Experiments on benchmark problems of sizes ranging from 5 to 10000 dimensions are conducted. The ESB variants perform better than line-search-based multi-start and simulated annealing techniques particularly when the function is noisy or the dimensionality of the problem is high. Several

hybrids under ESBF, involving simulated annealing, simplex and clustering methods are proposed.

In sum, the thesis has attempted to investigate the fundamental problem areas of computational learning and stochastic global optimization. Understanding the role of optimization as well as that of linear hypothesis classes in solving the two-class learning problem is given emphasis. The thesis illustrates the applicability of the results obtained in the above two areas through developing newer ANN models, ANN architecture construction and weight selection algorithms. The main strength of the work, however, lies in its potential in solving a broader range of problems in computational as well as management domains.

Related Publications

Sridev R. & Asim K Pal, *Learning with Linear Hypothesis Classes for Two Class Problems with Applications to ANN,* WPS-388, Working Paper Series, IIM Calcutta, 2000.

Sridev R. & Asim K Pal, *A Simplex-based Hybrid Algorithm for Unconstrained Function Optimization,* WPS-389, Working Paper Series, IIM Calcutta, 2000.

Sridev R. & Asim K Pal, *An Unified Framework for Stochastic Global Optimization,* WPS-390, Working Paper Series, IIM Calcutta, 2000.

Sridev R. & Asim K Pal, *An Elite Solution Based Framework for Stochastic Global Optimization,* WPS-391, Working Paper Series, IIM Calcutta, 2000.