Use of Bayesian Learning and Scaled Conjugate Gradient Method in ANN QSAR Models for HIV Proteases

Akmal Aulia[1], Sunil Kumar[2], Rajni Garg[1], A. Srinivas Reddy[2]

Introduction

- Recently, there have been numerous findings on the developments of new HIV drug candidates with various inhibitory activities.
- These activity variation correlates to structural changes among the drug candidates.
- The large number of structural properties (descriptors) associated with drug candidates makes it difficult for the traditional regression techniques to develop QSAR (Quantitative Structure Activity Relationship) models accurately.
- The use of machine learning techniques for structure-activity correlation has vastly increased over the past few years, due to the high accessibility of biological data and the increasing demand for more accurate and interpretable models for pharmaceutical development.

Materials and Methods

- Non-linear regression techniques are employed to analyze a large dataset of 335 compounds of HIV protease inhibitors (*Kempf et al.*).
- The dataset was studied using Hybrid GA-ANN (Genetic Algorithms-Artificial Neural Network) techniques to develop QSAR models, utilizing two different learning schemes, namely Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG).
- Each ligand (inhibitor or drug molecule) was described by means of physico-chemical and structural descriptors (features) which encode constitutional, electrostatic, geometrical, quantum and topological properties. The dataset used in this study consists of 277 descriptors and the associated EC_{50} activities.
- The capability of descriptors to address the variations in ligand(s) was linked to the predictive power of QSAR models.
- Combined information from these models helps in 'transforming data into information and information into knowledge' from *cheminformatics* point of view.

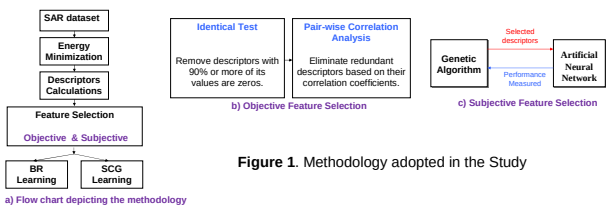

Figure 1. Methodology adopted in the Study

Bayesian Regularized Learning

- A common problem in performing regression via back propagation network is overfitting. As a result, the ANN would not yield accurate results for new inputs.
- Mackay [1] had proposed the Bayesian regularization learning scheme for optimizing the back propagation network model. Following Mackay's work, Foresee [2] came up with the idea to utilize Gauss-Newton approximation for computing the Hessian matrix.
- This computation can be implemented using the Levenberg-Marquadt algorithm, thus reducing the amount of computational overhead. In this study we have implemented Foresee's method with the help of Matlab's neural network toolbox [10].
- Originally, the optimization is done by minimizing the sum of squared errors:

$$F = E_D = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2$$

Where e_i represents the difference between the predicted activity and actual activity. However, in Mackay's work, the objective not only involves the errors generated by the outputs, but also the weights that reflect the connections between the nodes in the back propagation network model [1,2]. This performance index modification involves taking into account the sum of squares of the network weights:

$$F = \beta E_D + \alpha E_W$$

Scaled Conjugate Gradient Learning

- Unlike other ANN optimization methods, such as GD (gradient descent), CG (conjugate gradient), and BFGS (Broyden, Fletcher, Goldfarb, and Shanno quasi-Newton) method, the SCG (scaled conjugate gradient) is independent of user's specified parameters and avoids time-consuming line search [3].
- The CG and SCG are suitable class of optimization methods for large scale problems.
- In principle, the fast convergence of the SCG algorithm is due to avoidance of line-search on each iteration. Thus, in order to choose the search direction and scale the step size, SCG uses the quadratic approximations. However, for functions with non-positive definite Hessian matrices, having an iteration point distant from the preferred minimum will cause poor performance [3]. This can be compensated by using the Levenberg-Marquadt algorithm, thus the Hessian matrices can be kept positive definite.
- To determine the minimum point in the weight space (optimum ANN weights), SCG performs quadratic approximation of the error of point *w* in a neighborhood by:

$$E_{qw}(y) = E(w) + E'(w)^T y + \frac{1}{2} y^T E''(w) y$$

As described by Moller [3], the solution to the equation above are the critical points which minimizes it.

$$E'_{qw}(y) = E''(w)y + E'(w) = 0$$

Results

BR Neural Net Selected Descriptors	SCG Neural Net Selected Descriptors
Fractional PPSA (PPSA-3/TMSA) [Zefirov's PC]	Max partial charge for a H atom [Zefirov's PC]
HA dependent HDCA-2/SQRT(TMSA) [Zefirov's PC]	HA dependent HDCA-2/SQRT(TMSA) [Zefirov's PC]
Min 1-electron react. Index for a N atom	HACA-1/TMSA [Zefirov's PC]
Max net atomic charge for a N atom	XY Shadow
Randic index (order 3)	YZ Shadow
Complementary Information content (order 1)	Avg nucleoph. react. index for a C atom
Average Information content (order 2)	Average Information content (order 1)
Balaban index (_Property_)	Balaban index (_property_)

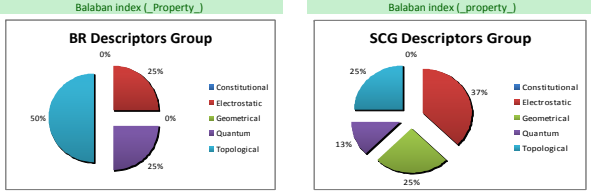

Figure 2. Classification of selected descriptors.

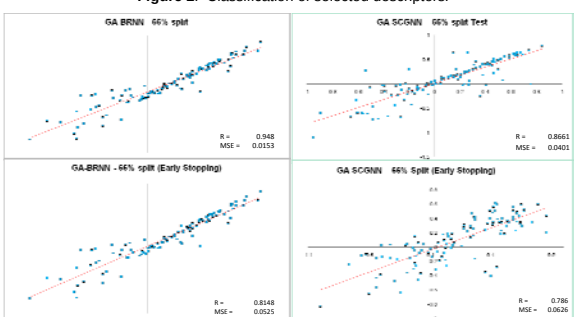

Figure 3 . Graphical depiction of Predicted (Y-axis) vs. Experimental activities (X-axis) along with their correlation coefficients and MSE

Discussions

The hybrid GA-ANN algorithm is a standard GA methodology utilizing ANN's performance as fitness function. Here, the early-stopped ANN has been implemented, in which the dataset was split into the training set (TSET) and the cross-validation set (CVSET). The splitting ratio was defined by Haykin as follow [6]:

$$r_{opt} = 1 - \frac{\sqrt{2W-1}-1}{2(W-1)}$$

where W is the total number of inputs; in this case, it will be the number of descriptors selected. Thus r_{opt} fraction of the dataset will be assigned as TSET, and the remainder as CVSET. Once the TSET and CVSET had been established, the ANN will be trained with the TSET, and CVSET will be simulated on a specified period of epochs (NN at n^{th} iteration) in order to seek an over fit. Once the whole ANN training sessions are done, the performance index (fitness) will be measured according to Guha as follows [5]:

$$fitness = TSET_{RMSE} + 0.5\,|\,TSET_{RMSE} - CVSET_{RMSE}\,|$$

- We have obtained 8 optimal descriptors for both GA-BRNN and GA-SCGNN.
- Although the number of selected descriptors are same, the composition of various components is different, but yielding the almost same prediction accuracy of the activity (in terms of their R & MSE). The role of geometrical descriptors seems to be taken by constitutional in case of GA-BRNN (Figure 3).
- BRANN gives considerably better prediction models than SCGANN with or without early stopping (Figure 3).
- But in case of early stopping, the differences in performances of both the models is small. Hence the early stopping seems to be crucial for SCGNN learning than BRNN.
- More experiments are in progress to under the strengths and week-nesses of the Bayesian over other learning schemes and role of early stopping.

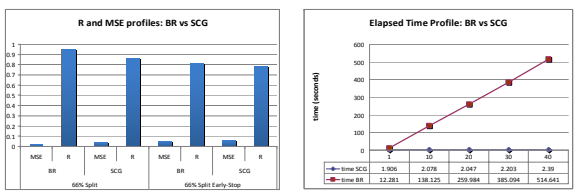

Figure 4 a) . Comparison of BR and SCG results based on the correlation coefficient (R) and the mean squared error (MSE). (b) Elapsed time profile based on epoch variations for 277 inputs (descriptors): BR vs SCG

Based on Figure 4b, although BRANN yields better model than SCGANN; SCGANN has the capability of delivering a reasonable model in a shorter amount of time when descriptors are more.

Acknowledgement

MATLAB 7.1, MATLAB Genetic Algorithm and Direct Search Toolbox, MATLAB Neural Network Toolbox. CODESSA 2.7.10.

References

1. D. J. C. MacKay, "Bayesian Interpolation", *Neural Computation*, vol. 4, pp. 415-447, 1992.
2. F. Dan Foresee and M.T. Hagan, "Gauss-Newton approximation to Bayesian learning", *International Conference on Neural Network*, vol. 3, pp. 1930-1935, 1997.
3. Martin F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning", *Neural Networks*, 4(6): pp. 962-969, 1993.
4. M. Fernandez and J. Caballero, "Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks", *J. Mol. Model.*, vol. 12, pp. 168-181, 2006.
5. R. Guha and Peter C. Jurs, "Development of QSAR models to predict and interpret the biological activity of Artemisinin analogues.", *J. Chem. Inf. Comput. Sci.*, vol. 44, pp. 1440-1449, 2004.
6. S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, New Jersey 1999.

[1]Computational Science Research Center; [2]Electrical and Computer Engineering Department, San Diego State University, San Diego, CA