財務管理論叢 第3卷 第2號 (1996.10)

A Multistrategy Learning System to Support Predictive Decision Making

Steven H. Kim and Heung Sik Oh*

ABSTRACT

The prediction of future demand is a vital task in managing business operations. To this end, traditional approaches often focused on statistical techniques such as exponential smoothing and moving average. The need for better accuracy has led to nonlinear techniques such as neural networks and case based reasoning. In addition, experimental design techniques such as orthogonal arrays may be used to assist in the formulation of an effective methodology.

This paper investigates a multistrategy approach involving neural nets, case based reasoning, and orthogonal arrays. Neural nets and case based reasoning are employed both separately and in combination, while orthoarrays are used to determine the best architecture for each approach. The comparative evaluation is performed in the context of an application relating to the prediction of Treasury notes.

* Graduate School of Management, Korea Advanced Institute of Science and Technology.

PURPOSE

The limitations of traditional approaches to the prediction of complex systems underscores the need for non-linear techniques. Multistrategy learning techniques such as neural networks and case based reasoning can provide superior performance in forecasting complex processes. Moreover, these techniques may be integrated into a synergistic system which performs better than the component techniques working in isolation.

MOTIVATION AND BACKGROUND

Experience with artificial intelligence applications, especially since the early 1980s, suggests the feasibility of a multistrategy approach to discovery and prediction for financial analysis. More specifically, knowledge-based systems may be employed to automate many of the decision making tasks involved in the use of particular techniques, to serve as a substrate to combine a multiplicity of methodologies, and to improve system performance by learning to identify the utility of different combinations of techniques. The advantages of combining multiple techniques to yield synergism for discovery and prediction have been recognized in the past (Kaufman et, al., 1991; etc.).

A versatile approach to self-organization lies in neural networks (Anderson and Rosenfeld, 1988; Grossberg, 1974, 1976; Haken, 1988; Hebb, 1949; Hopfield, 1982; Kohonen, 1984; Rosenblatt, 1962; Rumelhart et al., 1986). Neural nets are characterized by learning capability, the ability to improve performance over time. A closely related feature is that of generalization, relating to the recognition of new objects which are similar but not identical to previous ones. An additional characteristic relates to graceful degradation: the network fails gradually rather than catastrophically when it suffers partial damage.

To date, however, artificial networks have been burdened with a major limitation: protracted training periods. Hundreds or thousands of trials are usually required for satisfactory performance in various tasks. The time and effort required for training have hindered their widespread application to practical domains (Kim, 1994; Shibazaki and Kim, 1992). To fully exploit the promise of neural nets by emulating the real-time responsiveness of biological systems, training time must be reduced dramatically, by orders of magnitude. Performance improvement of such magnitude will not likely materialize from a simple tweaking of algorithms or their parameters. A more drastic re-evaluation and improvement in technique are indicated.

In neural networks, the knowledge is encoded implicitly in the form of link weights and threshold values; the network may perform admirably, but its newly acquired knowledge remains in a black box. The fusion of declarative and neural knowledge, however, offers the potential for making explicit part of the knowledge encoded in a network.

The use of explicit knowledge allows for explanation and justification for the benefit of other entities, including an interested human observer. Examples of such high-level representation, also called the knowledge level, lies with case reasoning. A sophisticated learning system should provide for the fusion of both implicit and explicit methods of knowledge representation. In this way, it can build on the respective advantages of disparate techniques.

METHODOLOGY

Neural Networks (NN).

The most common neural network methodology employs the backpropagation algorithm. This approach involves a layered, feedforward network structure with fully interconnected nodes from one layer to the next. The learning technique involves backward propagation of errors to aid in updating internode weights. As a model of biological systems, artificial neural networks have learning capabilities which can be applied to the task of prediction.

Case Based Reasoning (CBR).

Case based reasoning has been investigated extensively over the past decade. The key idea in CBR is to determine the nearest neighbors in a case base, and to use the neighbors to gain insight into unknown attributes of the target case. Techniques such as cluster analysis may be used to enhance the performance of case reasoning systems (Kim and Novick, 1993).

Orthogonal Arrays (OA).

Orthoarrays are useful for designing experiments and determining key effects in a complex situation. Orthoarrays may be used as a primary technique for pattern recognition to identify the effective values of various parameters in designing a system with nonlinear attributes.

However, a more common application of an orthoarray is as a complementary technique. For instance, an orthoarray may be used to determine the best combination of input vector size and number of neighbors to consider in a case reasoning system.

Table 1. Characteristic of the data set: monthly data for Treasury notes of 10 years maturity.

	Learning	Testing
Start date	Jan. 1960	Mar. 1981
End date	Feb. 1981	Dec. 1992
Numerosity	254	142

Table 2. Candidate configurations for the integrated models (NN_CBR). Each of the 3 parameters (NN architecture, input vector size to CBR, and number of neighbors for CBR) exhibits 3 possibilities.

Level	Factor		
	Architecture of NN	Input Vector Size (D) for CBR	# of Neighbors (J) for CBR
1	2 x 3 x 1	2	2
2	3 x 3 x 1	6	6
3	4 x 3 x 1	10	10

Neural networks, case based reasoning, and orthoarrays may be used to uncover previously unknown patterns in a database. All these methods are employed in the case study described in the next section.

CASE STUDY

The case study involves a comparative evaluation of learning techniques for predicting a complex process. More specifically, neural networks and case reasoning are employed – both individually and jointly – in the complex task of forecasting U.S. Treasury Notes of 10 years maturity. The details of the data set are given in Table 1.

Figure 1 shows the general architecture for the neural network models used in isolation. Of the architectures evaluated, the best performance resulted from a 3 x 3 x 1 configuration: 3 input nodes, 3 hidden nodes, and 1 output node.

Figure 2 depicts the general architecture for the case based reasoning models. As with the neural network approach, the input variables were a sequence of current and lagged variables: $X_t, X_{t-1}, \cdots, X_{t-k}$. Of the configurations tested, the best model resulted from D = 10 input variables and J = 10 neighbors.

The general architecture for the integrated methodology is presented in Figure 3. All the input variables enter the CBR module as well as the NN module.

To determine the best integrated model, the architecture was varied along 3 dimensions as shown in Table 2. The results of the orthoarray experiments are shown in Table 3; the implication is that the second NN configuration (3 x 3 x 1), second input vector size (D = 6), and third neighborhood size (J = 10) represent the best integrated model.

The performance of the best CBR model is shown in Figure 5; and the best NN model in Figure 6. Both models work fairly well. Since the integrated model performed even better than the two isolated models, the corresponding chart is not shown.

Figure 7 plots the neighbors from the best NN, CBR, and integrated models. The difference in performance is highlighted by the mean square error of forecast in Table 4. The data indicate that the integrated model (NN_CBR) outperforms the CBR model, which in turn supercedes the NN model. In addition, a set of pairwise t-tests in Table 5 indicate that the differences are significant.

Figure 1. General architecture for neural network (NN) models in isolation. The following architectures were tested: 2 x 3 x 1, 3 x 3 x 1, and 4 x 3 x 1. Among these, the best performance resulted from the 3 x 3 x 1 architecture.

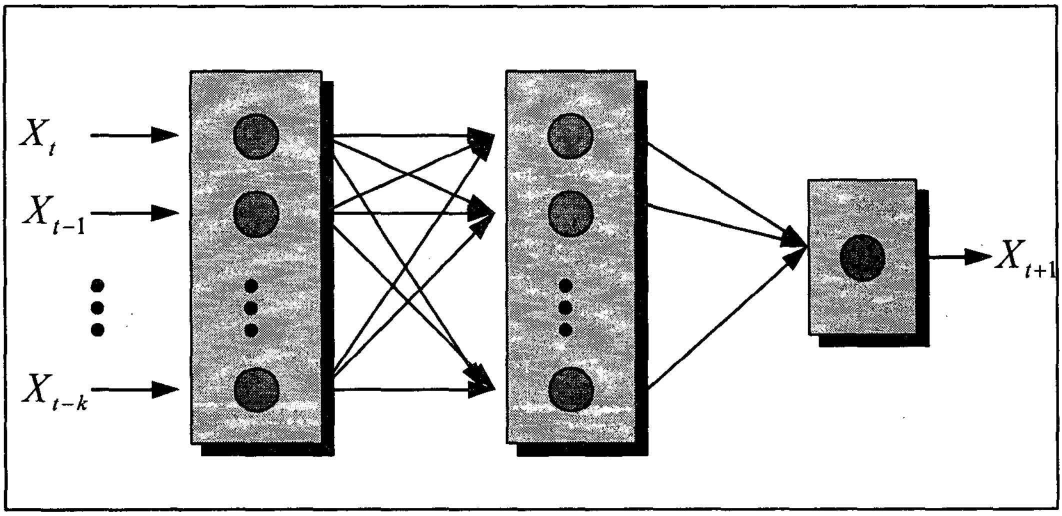

Figure 2. General architecture for case based reasoning (CBR) models. The following configuration were tested: input vector size of D = 2,6,10; and number of neighbors J = 2,6,10. The best performance resulted from D = 10 and J = 10.

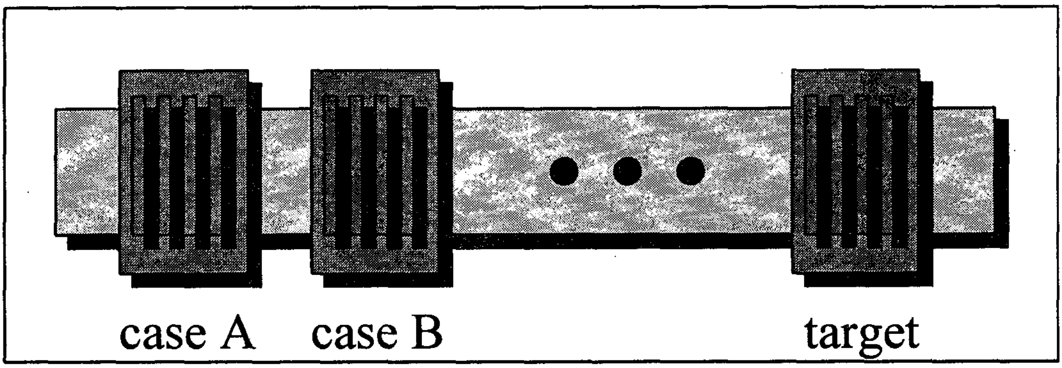

Figure 3. General architecture for the integrated methodology (NN_CBR).

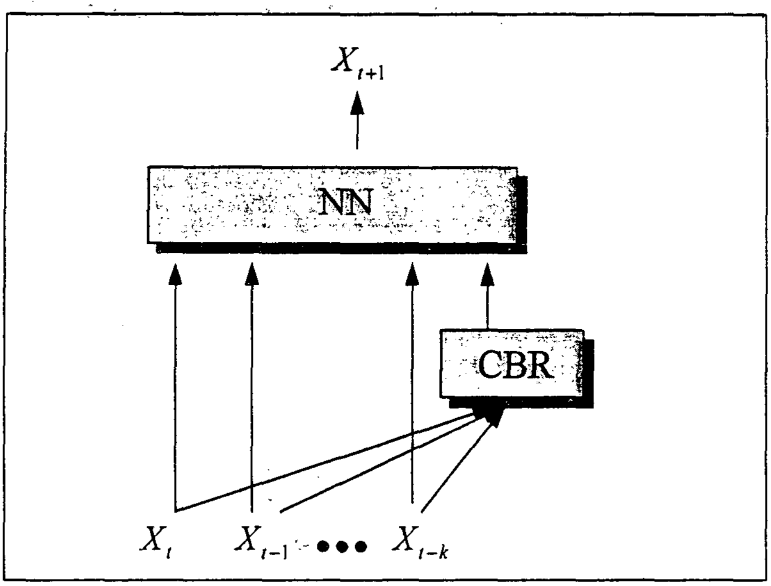

Figure 4. Effect of each factor in the orthogonal array experiments for the integrated model.

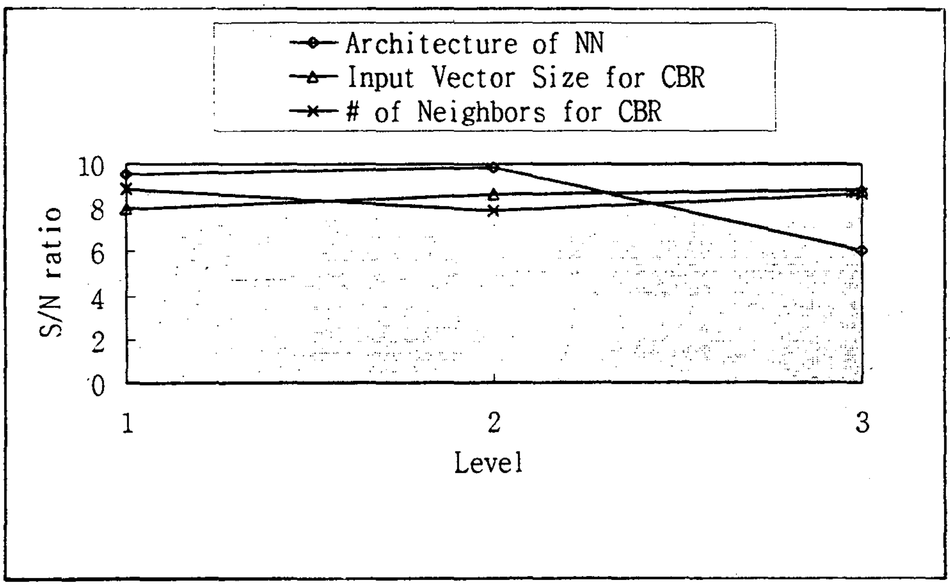

Figure 5. Performance of the best CBR model.

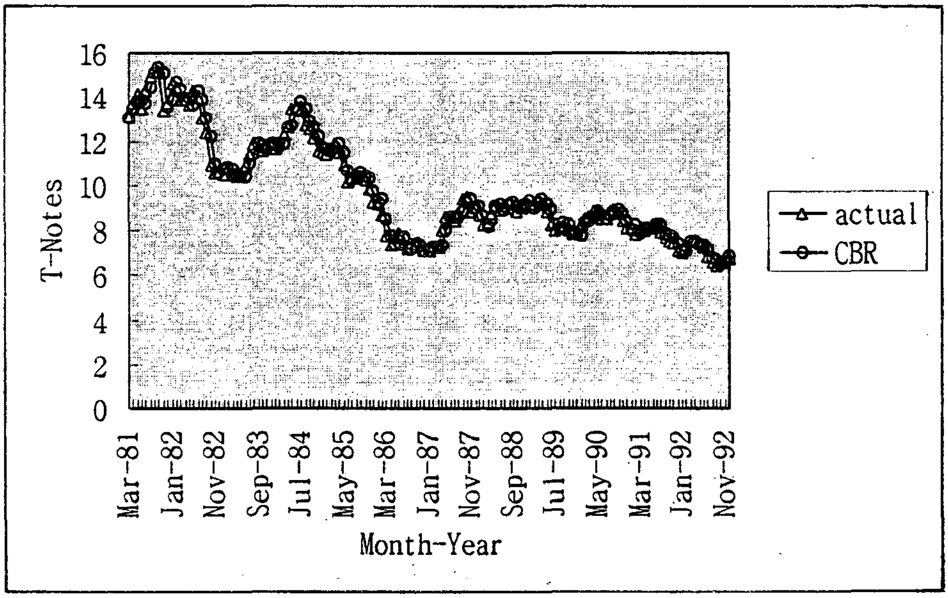

Figure 6. Performance of the best NN model.

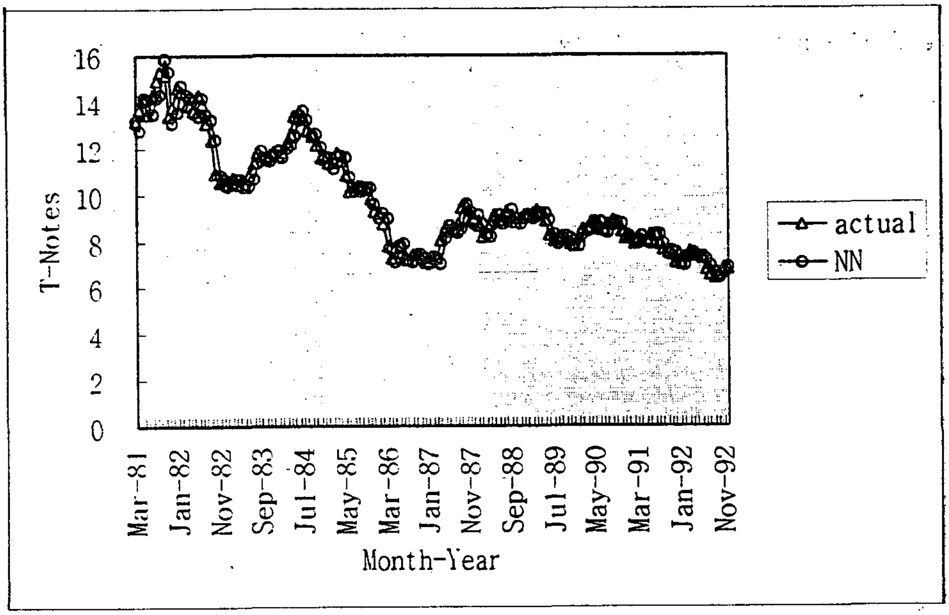

Figure 7. Residuals from forecasts by the best NN model, best CBR model, and best integrated model.

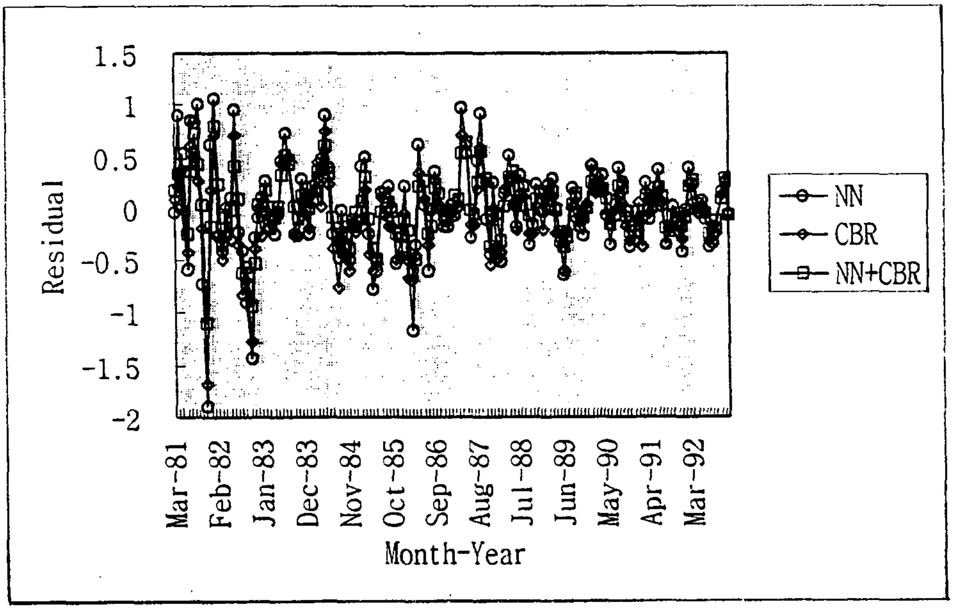

Table 3. Results of the orthoarray experiments for the candidate models defined in Table 1.

Eep #	Factor			Observation
	Architecture of NN	Input Vector Size for CBR	# of Neighbors for CBR	
1	1	1	1	9.9482
2	1	2	2	8.9415
3	1	3	3	9.5468
4	2	1	2	8.4043
5	**2**	**2**	**3**	**10.6198**
6	2	3	1	10.4528
7	3	1	3	5.5269
8	3	2	1	6.216
9	3	3	2	6.2525

Table 4. Performance of the various models based on forecast error.

Model 1	MSE
NN	0.2011
CBR	0.1329
NN_CBR	0.0867

Table 5. Pairwise t-tests for the models in Table 3.

Models	t-statistic	p-value
NN vs. CBR	4.07	0.000
NN vs. NN_CBR	4.84	0.000
CBR vs. NN_CBR	3.32	0.001

CONCLUSION

The results of this study reveal that CBR can perform better than neural networks. Moreover, the integrated model comprising both case reasoning and a neural net is significantly more accurate than either the CBR or the neural network working separately.

A knowledge system for predicting complex processes is applicable to the fields of finance, investment, banking, and economics as well as science and technology. The integrated methodology for knowledge discovery and prediction described in this paper will pertain to many areas rich in databases, both within the financial community and without.

REFERENCES

Anderson, J. A. and Rosenfeld, E., eds. Neurocomputing. Cambridge, MA: MIT Press, (1988).

Grossberg, S., "Adaptive Pattern Classification and Universal Recoding: Part I. Parallel Development and Coding of Neural Feature Detectors," *Biological Cybernetics* 23(3), (1976), 121-134.

Grossberg, S., "Classical and Instrumental Learning by Neural Networks," *Progress in Theoretical Biology* 3. NY: Academic Press, (1974), 51-141.

Haken, H., ed. *Neural and Synergetic Computers.* New York: Springer-Verlag, (1988).

Hebb, D. O., *The Organization of Behavior,* NY: Wiley, (1949).

Hopfield, J. J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Nat. Acad. Science USA* 79(8), (1982), 2554-2558.

Kaufman, K. A., R. S. Michalski, L. Kerschberg, "Mining for Knowledge in Databases: Goals and General Description of the INLEN System. In G. Piatetsky-Shapiro and W. J. Frawley, eds.," *Knowledge Discovery in Database*, Cambridge, MA: AAAI/MIT Press, (1991), 449-462.

Kim, S. H., *Essence of Creativity,* New York: Oxford Univ. Press, (1990).

Kim, S. H., *Learning and Coordination.* Dordrecht, Netherlands: Kluwer, (1994).

Kim, S. H. and M. B. Novick, "Using Clustering Techniques to Support Case Reasoning," *International J. of Computer Applications in Technology* 6(2/3), (1993), 57-73.

Kohonen, T., *Self-Organization and Associative Memory.* NY: Springer-Verlag, (1984).

Rosenblatt, F., *Principles of Neurodynamics.* NY: Spartan, (1962).

Rumelhart, D. E., G. E. Hinton and R. J. Williams, "Learning Internal Representations by Back Propagation. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group," *Parallel Distributed Processing* 1: Foundations, Cambridge, MA: M.I.T. Press, (1986), 318-362.

Shibazaki, H. and S. H. Kim, Learning Systems for Manufacturing Automation: Integrating Explicit and Implicit Knowledge. Presented at International Conference on the Manufacturing Science and Technology of the Future, Stockholm, Sweden, June 1989, Also in *Robotics and Computer-Integrated Manufacturing* 9, (1992).