IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITP4490

Artificial Intelligence in an Ecological Application

M. L. Hines and J. Pang
University of Missouri-Kansas City, Kansas City, MO 64110, Ph: (816) 235-1495, Fax: (816) 235-5159, hines@cstp.umkc.edu, jpang@cstp.umkc.edu

C. Butterfield
Chadron State College, Chadron, NE 69337, Ph: (308) 432-6274, Fax: (308) 432-6369, cbutterfield@csc.edu

I. INTRODUCTION

The Range Ecology group of the Department of Agronomy at the University of Nebraska, Lincoln, has been investigating the natural re-vegetation of an area previously grazed by cattle. The primary focus has been how the plants will change and how long re-vegetation will take. For the purpose of that research, the area has been fenced for fifteen years since 1981, so cattle cannot get into the area. During the research, fifteen-year plant community data were recorded. A tool was needed to enable predictive analysis.

This research evaluated two artificial intelligence approaches for the predication of ecological plant community succession: artificial neural networks and knowledge based systems, two of the most widely used and commercially successful applications of artificial intelligence. A base prototype for predicting plant community succession model was then built.

Section two overviews the two AI approaches evaluated. Section three describes the ecological application that prompted this work and the rationale to determine which AI approach to use. Section four presents a high-level overview of the prototype developed and the results from training. Section five outlines the work yet to be done and summarizes the potential impact of this research.

II. KNOWLEDGE BASED SYSTEMS VS. ARTIFICAL NEURAL NETWORKS

The main goal of AI is to build systems that exhibit intelligent behavior and perform complex tasks with a level of competence equivalent or superior to the level currently exhibited by a human expert [12]. Artificial neural networks have emerged from the connectionist approach, and knowledge based systems have emerged from the symbolic approach.

A knowledge based system, also known as an expert system, is a computer program that uses knowledge and inference procedures to solve problems. To solve expert-level problems, expert systems need access to a substantial domain knowledge base, need to exploit one or more reasoning mechanisms to apply knowledge to the problems, and need a mechanism for explaining to users what they have done [17]. Domain knowledge is represented as a set of production rules [17].

An expert system usually needs to interact with the user, the environment and other systems such as databases. The input/output interfaces are usually displaying output and receiving input either from the user or directly from other devices.

Knowledge based systems have been used in vegetation science. Loh and Hsieh [8] linked conventional simulation approaches to a raster-based GIS package and a rule-based expert system shell to model secondary woody plant succession at the landscape level.

The architecture of artificial neural networks (ANNs) is a model of the brain's cognitive process, where the basic units are artificial neurons. ANNs are trained by "catching" the knowledge embedded in examples. Once trained, the ANNs can generate correct output when new similar input data are given.

One approach to make a network intelligent is to store the "knowledge" in the "synapses," the weights of the connections between the neurons. The network acquires knowledge during "training". Input-output pattern associations are presented to the neural network in sequence. The network adjusts its weights to capture the knowledge embedded in these pattern pairs. Once the knowledge is present in the synaptic weights of the network, presenting a pattern for input to the network will produce the correct output.

In many neural network models, learning takes the form of supervised training. The input-output pattern pairs are presented one by one to the neural network and the actual output is compared with the desired output. Two-layer networks have only two layers, an input layer, and an output layer. The relationship between input and output patterns is linear. Two-layer networks are limited in that they are not able to learn nonlinear relationships between input and output patterns.

An additional layer of neurons is added between the input and output layers, and the activation function is defined for the hidden-layer neurons to model a nonlinear function.

This additional layer is called a "hidden layer" since it does not interact directly with the outside. The input layer is made up of the neurons receiving inputs from the outside. The output layer is made up of the neurons that receive inputs from hidden-layer neurons and generate the outputs.

The feedforward, backpropagation network is a multilayer

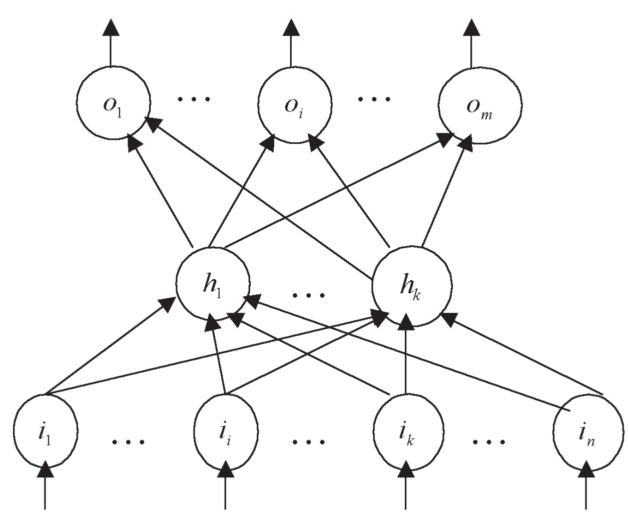

Fig.1. Diagram of backpropagation topology

network with supervised training. A schematic three-layer backpropagation ANN model is shown in Figure 1.

The backpropagation model presented here consists of three layers of neurons: an input layer, a hidden layer, and an output layer. Two layers of synaptic weights exist, one between the input layer and hidden layer, and the other one between the hidden layer and output layer. Input data are fed to the processing neurons at the input layer, signals are propagated through the hidden layer, and the results of network processing are produced at the output layer.

Supervised training requires that a set of "good" input-output pattern pairs be used during the training operation. During the feedforward operation, the input neurons distribute inputs in parallel simultaneously to the hidden layer neurons. The output of hidden units is calculated by applying the activation function to the net input.

The output is calculated in the same way. When the actual outputs corresponding to the input pattern of the network are generated, the errors, which are the difference between the forecasted outputs and the actual outputs, are calculated. If the error is not zero, corrections are made in the weights in proportion to this error. The theory of the backpropagation method involves making the corrections to the weights from the last-but-one layer to the last layer first, then using the calculations involved in these corrections as the basis for calculating the corrections for the next layer back … until the input layer is reached.

The knowledge of a neural network is stored in the weights of the connections between the neurons. That knowledge is captured during the training phase, where the feedforward output state calculation is combined with backward error propagation and weight adjustment calculations representing the network's learning, or training. The goal of the training process is to minimize the difference between forecasted and actual output value over all training patterns.

Artificial neural networks have been used in ecological modelling. Tan and Smeins [19] built a feedforward, backpropagation neural network model to predict percent composition of two plants from knowledge of present climatic factors and species cover. The resulting trained ANN is capable of forecasting accurately up to 4 years into the future. The result indicates a potential usefulness of neural network technology for non-mechanistic modeling in ecological research and management.

III. AN ECOLOGICAL APPLICATION AND MODEL

Research conducted by the Range Ecology group of the Department of Agronomy at the University of Nebraska, Lincoln, investigated whether an area, which was grazed by cattle before, will cover over with vegetation on its own. If it will, how do the plants change and how long will it take. For the purpose of the research, the area has been fenced for fifteen years so cattle cannot re-graze the area. During the study, plant community data were recorded.

This research was conducted from 1981 to 1995 at the Nature Conservancy's Niobrara Valley Preserve. The 12 hectares site was around a water tank located in a pasture formerly grazed by cattle. In 1981 the entire pasture containing the study site was excluded from grazing to form a study pasture.

A grid system consisting of 272 points, with 15 meters between points, was established in the research site. Each point on the ground is a 1.5m length of electrical conduit permanently marked 0.5m from the bottom. Each piece of conduit was driven into the sand until the permanent mark was level with the current sand surface level. In succeeding years the sand level was measured at each point as either erosion or deposition, to the nearest 5mm, then the point reset so the mark was at the current sand surface level. In 1989 a level circuit was run on all established points to determine the elevation of each point. Sand erosion and deposition data was then added or subtracted for each preceding year to determine the 1980 elevation of the study site.

Each differing plant patch within the blowout is mapped using the grid points as a guide. Within each of the mapped patches basal cover is obtained using a 10-point sampling frame. All species present within each mapped patch are also noted along with the dominant plant or plants. Aerial photographs of the blowout are taken on an annual basis. Starting in 1994 individual species frequency data was obtained within each mapped patch utilizing a 0.10 m^2 square frame. Frames were located by a random pace method utilizing the points on the ground to keep within the confines of the patch and to avoid any potential edge effects.

In 1992 and 1993 soil samples were collected for soil chemistry data. Within the differing mapped patches numbered conduit were randomly selected. Individual soil samples were then obtained four feet from each conduit, in each of the 4 cardinal directions. Samples were analyzed for Kjeldahl nitrogen, pH, phosphorous, potassium, and organic matter.

The research region then consists of 272 cells with each cell representing a 15 m x 15 m area, and associated with a plant community class. Determination of the appropriateness of either an expert system or an artificial neural network for predicting successions of plant communities was evaluated. The expert system approach was evaluated first.

The first step, knowledge acquisition, was performed by interviewing the domain expert and reviewing ecological journals about vegetation dynamics such as [1], [4], [5], [7], [9], [10], [11], [13], [14], [15], [16], [18]. Several interviews were conducted with the University of Nebraska research technologist to elicit knowledge related to the domain problem, such as the basic progression of plant communities, sand movement, spatial statistics, and soil chemistry. This knowledge was used to formulate knowledge rules, which upon further examination revealed a number of anomalies. It was difficult for the expert to explain these anomalies.

The impact of expert systems on ecological theory depends on the degree to which "deep knowledge" (i.e., knowledge based on first principle rather than on more empirical rules) is used in formulating knowledge bases [13]. The knowledge associated with this particular research domain was based on empirical rules, not deep knowledge. Thus a knowledge based system approach was not suitable for this particular research. The next step was to evaluate the suitability of an artificial neural network approach. There were three necessary conditions to be considered for using an artificial neural network for this problem.

The first necessary condition was that the inputs to the network contain sufficient information relating to the output. In the class level spatial statistics, the area and the percent composition of each plant community class for that year were available. The precipitation of each study year was available too. The information relating to the plant community successions was sufficient.

The second necessary condition for using an artificial neural network approach was that there exist relationships between inputs and outputs. The succession is a directional cumulative change in the types of plant species that occupy a given area through time. The species occupation in a year strongly influences the species composition in the coming year. The climatic factors also play an important role in the growth of plant communities. So, the relationships between the inputs and outputs exist.

Fifteen years continuous data recorded about plant community changes satisfies the third necessary condition which requires a sufficient input-output pattern of facts to train the network although it later turned out that not all of this data was readily available.

Previous research successfully used artificial neural networks in an ecological modelling application to predict the percent compositions of two plants up to four years into the future [19]. It is possible to take advantage of the ability of a neural network to learn by examples, store the knowledge into the weights, recognize and generalize from patterns contained in the ecological data. So, it is reasonable to build a neural network model to predict the percent composition changes of the nine plant communities.

A model was formulated to serve as the basis for the plantANN model. This consisted of three primary steps: determine the outputs, determine the inputs, choose the neural network topology and define its parameters.

The outputs of the plantAnn contain numbers determining the percentage of the study area for all plant communities. One goal of this research was to build a base prototype to predict the plant community successions by predicting what plant communities will appear in the next coming year, and their areas. Percentage of study area of each plant community occupied is examined instead of the actual area.

The next step is to choose as many input factors as possible that might be related to the plant community composition changes. The possible relevant input factors were determined and included in the model.

1. Current year plant community compositions*:* The obvious input factors are the current year compositions of plant communities. These input factors would be the values of percentage of study area of each class. If a class does not appear in the current year, a zero value should be assigned to that input.
2. Previous two years plant community compositions*:* Changes in plant community compositions two years earlier may also have predictive power. Using previous two year data should increase the number of input factors, but will give a more effective neural net. These input factors would be the percentage of study area of each class. If a class did not appear in that year, a zero value should be assigned to that input.
3. Precipitation*:* The growth of plants is influenced by the precipitation. The average precipitation of the current year is included in the input factors.
4. Previous two year precipitation*:* Since the previous two year plant community compositions were included in the input factors, the previous two year average precipitation data should also be included.
5. Temperature*:* Temperature influences the growth of plants also. Currently, no data about the temperature is available. If the temperature data becomes available, it should be included into the input vector.

Once the expected outputs and input factors were chosen, the next step was to choose a neural network topology and define its parameters. To capture the nonlinear relationship between input and output patterns, a multilayer network was needed. A three-layer network, containing one hidden layer capable of capturing nonlinear relationships to the required level of complexity, was chosen.

Once the topology was chosen, its parameters were defined. In this research, the input layer consists of thirty neurons, containing the data for the current year precipitation and plant community compositions, and previous two years precipitation and plant community compositions. The output layer consists of nine neurons, which output the prediction for the coming year's percent composition of plant communities.

Determination of the size of the hidden layer is generally by trial-and-error. The heuristic rule is to choose a number between the number of input units and output units. Initially the number of hidden layer units was chosen as twenty-five.

The learning rate cannot be determined optimally before the application is run, but 0.5 is a reasonable first entry [3]. The momentum term, which is the amount the previous weight change affects the current weight change, was defined as 0.2. The tolerance of this network was defined as 0.1, which means that any prediction within 10 percent of the desired result was acceptable.

IV. MODEL IMPLEMENTATION AND USE

The artificial neural network model with a feedforward, backpropagation neural network algorithm was modified from an existing implementation [3] using Visual C++. The prototype consists of five classes: the vector class, the matrix class, the vector pair class, the neural network class, and the backpropagation class.

A neural network consists of several layers, each layer consisting of neurons represented as a vector. Floating-point vectors are used as a "worst case" that can handle any type of vector that needs to be represented. The member functions of the vector class consist of the constructor to create objects, the destructor to destroy objects, and a host of arithmetic operators.

The connections between neurons are represented as matrices with two numbers representing the number of rows and columns. In the matrix class, several constructors are provided to construct the matrix object according to different purposes. A destructor destroys objects, and standard matrix arithmetic functions were implemented.

The neural network captures knowledge by learning association between input-output patterns. The vector pair is represented by two vector objects. Methods are provided for assignment and testing for equivalence. Synapses "is-a" matrix, which means that synapses is derived from an object matrix.

Besides using vectors to represent neurons in different layers and matrices to represent synapses connecting neurons, a neural network must be able to learn examples. It also needs to be tested from another set of facts, which are not used in the training phase, and to be run on a set of inputs when it is trained and tested. The neural network class declares the methods for train, test, and run. The train method runs a set of presented facts until the desired tolerance is reached. The test method runs on a set of input facts to compare the actual outputs with desired outputs and returns a value indicating the percentage correct of test. The run method will run a set of inputs and generate outputs from network.

The backpropagation class is derived from the neural network class and inherits the properties of that class. Present in the backpropagation class are the data and methods unique to backpropagation.

The neural network implemented needed to be trained first, so it could perform prediction. The training and testing data were prepared from the class level statistics and precipitation data. It included annual information about precipitation and plant community composition for: bare sand, blowout grass, blowout grass/sand muhly, perennial grass/sand muhly, perennial grass, new sandhill prairie, annual grass, annual/perennial grass, and shrubs.

Once all data were gathered, it was used to train the network. Usually, when a network is trained, the network is ready to be tested. The test data stored in the test file is sent to the network, then the actual outputs are compared with the desired output. Only

Table 1

Plant Community Type	Forecasted (%)	Actual (%)	Difference (%)
Perennial grass/sand muhly	0.00	0.00	0.00
Blowout grass/sand muhly	1.00	0.00	-1.00
New sandhill prairie	2.00	0.00	-2.00
Bare sand	6.00	11.68	5.68
Blowout grass	31.00	27.21	-3.79
Perennial grass	38.20	29.84	-8.36
Annual	11.00	19.83	8.83
Annual/perennial grass	4.30	7.90	3.60
Shrub	6.00	3.53	-2.47

the data from 1981 to 1989 were available. Since the data were not enough for training and testing, testing data were not separated from training data. As more data becomes available, testing data needs to be separated from training data.

1985's, 1984's, 1983's precipitation and plant composition data were used as input data to predict 1986's plant compositions (table 1).

The average difference between the forecasted results and actual results was 0.039, which falls into the tolerance range of 0.1.

There are some possible factors causing the differences between the forecasted results and actual results of some communities. One factor is the inconsistency of the data collection process. Another factor for some communities is the amount of data used in the training operation may not have been sufficient. Eight years data from 1981 to 1989 were constructed to make five training input-output patterns and one running pattern. More data is needed for the training operation. Temperature also plays an important role in the growth of plants but was not available.

V. SUMMARY

The Range Ecology group at the University of Nebraska, was investigating whether an area grazed by cattle, will cover over with vegetation on its own. If it will, how do the plants change and how long will it take. Prediction of plant community successions arose from this research. The purpose of this research was to evaluate the knowledge based and artificial neural network approaches in this ecological application and build a base prototype for predicting plant community succession.

Initially, knowledge based systems and artificial neural network technologies were reviewed and evaluated with regard to this ecological application. A knowledge based system approach was not suitable for this particular problem because of the lack of deep knowledge. Artificial neural networks use supervised training to capture embedded knowledge from the facts used in training operations. It was determined reasonable to use artificial neural network technology to solve this problem.

A base prototype model for predicting plant community successions was built. The outputs of the neural network model contained the percent composition of plant communities. The inputs of the neural network model contained the current year precipitation and plant community compositions and previous two years precipitation and plant community compositions. The topology of the neural network was a feedforward, backpropagation, supervised training network. The available data from 1981 to 1989 were used to form five training data sets and one running data. The prediction result fell within the tolerance range for being successful even though there were differences. There are several factors potentially causing the difference between the predicted result and actual result of some plant communities: the inconsistency of the data collection process, insufficient training data, lack of temperature data. The research indicates artificial neural networks are a potential technology for predicting plant community succession.

VI. REFERENCES

1. Balzter, H., Braun, P.W., and Kohler, W. Cellular automata models for vegetation dynamics. *Ecological Modelling* 107 (1998), 113-125
2. Bharath, R., and Drosen, J. *Neural Network Computing*. Windcrest, New York. 1994.
3. Blum, A. *NEURAL NETWORKS IN C++; an Object-Oriented Framework for Building Connectionist Systems*. John Wiley & Sons, Inc. New York. 1992.
4. Colasanti, R. L., and Grime, J. P. Resource dynamics and vegetation processes: A deterministic model using two-dimensional cellular automata. *Functional Ecology* 7 (1993), 169-176.
5. Crawley, M. J., and May, R. M. Population dynamics and plant community structure: Competition between annuals and perennials. *J. Theor. Bilol*. 125 (1987), 475-489.
6. Ignizio, J.P. *Introduction to Expert Systems: The Development and Implementation of Rule-Based Expert Systems*. McGraw-Hill, Inc., New York. 1991.
7. Itami, R. M. Simulating spatial dynamics: cellular automata theory. *Landscape and Urban Planning* 30 (1994), 27-47.
8. Loh, D.K., and Hsieh, Y. – T.C., Incorporating rule-based reasoning in the spatial modeling of succession in a savanna landscape. *AI Appl*. 9 (1995), 29-40.
9. McCook, L. J. Understanding ecological community succession: Causal models and theories, a review. *Vegetatio* 110 (1994), 115-147.
10. Moore, A.D., and Noble, I.R. Automatic model simplification: The generation of replacement sequences and their use in vegetation modelling. *Ecological Modelling* 70 (1993), 137-157.
11. Moore, A.D., and Noble, I.R. An individualistic model of vegetation stand dynamics. *Journal of Environmental Management* 31 (1990), 61-81.
12. Nikolopoulos, C. *Expert Systems: Introduction to First and Second Generation and Hybrid Knowledge Based Systems*. Marcel Dekker, Inc. New York. 1997.
13. Noble, I.R. The role of expert systems in vegetation science. *Vegetatio* 69 (1987), 115-121.
14. Noble, I.R., and Habiba, G. A functional classification for predicting the dynamics of landscapes. *Journal of Vegetation Science* 7 (1996), 329-336.
15. Noble, I.R., and Slatyer, R.O. The use of vital attributes to predict successional changes in plant communities subject to recurrent disturbances. *Vegetatio* 43 (1980), 5-21.
16. Patoucheas, D.P., and Stamou, G. Non homogeneous Markovian models in ecological modelling: A study of zoobenthos dynamics in Thermaikos Gulf, Greece. *Ecological Modelling* 66 (1993), 197-215.
17. Rich E., and Knight, K. *Artificial Intelligence*. 2d ed. McGraw-Hill, Inc. New York. 1990.
18. Salski, A. Fuzzy knowledge-based models in ecological research. *Ecological Modelling* 63 (1992), 103-112.
19. Tan, S.S., and Smeins, F.E. Predicting grassland community changes with an artificial neural network model. *Ecological Modelling* 84 (1996), 91-97.