Title
Unit 188 - Artificial Neural Networks for Spatial Data Analysis

Authors
188, CC in GIScience
Gopal, Sucharita

Publication Date
2000

Peer reviewed

Unit 188 - Artificial Neural Networks for Spatial Data Analysis

Written by Sucharita Gopal
Department of Geography and Centre for Remote Sensing
Boston University, Boston MA 02215

Advanced Organizer

Topics covered in this unit

Intended Learning Outcomes

After learning the material in this unit, students should be able to:

- Define ANN and describe different types and some applications of ANN
- Explain the applications of ANN in geography and spatial analysis
- Explain the differences between ANN and AI, and between ANN and statistics
- Demonstrate a broad understanding of methodology in using ANN
- Apply a supervise ANN model in a classification problem
- Apply a supervised ANN in a function estimation problem

Full Table of Contents

Metadata and Revision History

Unit 188 - Artificial Neural Networks for Spatial Data Analysis

1. Introduction

1.1. What are Artificial Neural Networks (ANN)?

- provide the potential of an alternative information processing paradigm that involves
  - large interconnected networks of processing units (PE)
  - units relatively simple and typically non-linear
  - units connected to each other by communication channels or "connections"
  - connections carry numeric (as opposed to symbolic) data; encoded by any of various means
  - units operate only on their local data and on the inputs they receive via the connections

1.2. Some Definitions of ANN

- According to the DARPA Neural Network Study (1988, AFCEA International Press, p. 60):
  - *a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes*.

- According to Haykin, S. (1994),Neural Networks: A Comprehensive Foundation, NY: Macmillan, p. 2:
  - *A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain processor in two respects:*
    1. Knowledge is acquired by the network through a learning process.
    2. Interneuron connection strengths known as synaptic weights are used to store the knowledge.

1.3. Brief History of ANN

- ANN were inspired by models of biological neural networks since much of the motivation came from the desire to produce artificial systems capable of sophisticated, perhaps "intelligent", computations similar to those that the human brain routinely performs, and thereby possibly to enhance our understanding of the human brain.

1.4. Applications of ANN

- ANN is a multi-disciplinary field and as such its applications are numerous including

- finance
- industry
- agriculture
- business
- physics
- statistics
- cognitive science
- neuroscience
- weather forecasting
- computer science and engineering
- spatial analysis and geography

1.5. Differences between ANN and AI approaches:

- Several features distinguish this paradigm from conventional computing and traditional artificial intelligence approaches. In ANN
  - information processing is inherently parallel.
  - knowledge distributed throughout the system
  - ANNs are extremely fault tolerant
  - adaptive model free function estimation, non-algorithmic strategy

1.6. ANN in Apatial Analysis and Geography

- Fischer (1992) outlines the role of ANN in both exploratory and explanatory modeling.
- Key candidate application areas in *exploratory* geographic information processing are considered to include:
  - exploratory spatial data and image analysis (pattern detection and completion, classification of very large data sets), especially in remote sensing and data rich GIS environments (Carpenter et al., 1997)
  - Regional taxonomy including functional problems and homogenous problems (See Openshaw, 1993)
- Key candidate application areas in *explanatory* geographic information processing
  - spatial interaction modeling including spatial interaction analysis and choice analysis (e.g. Fischer and Gopal, 1995)
  - optimization problems such as classical traveling salesman problem and shortest-path problem in networks (Hopfield and Tank, 1985)
  - space-time statistical modeling (Gopal and Woodcock, 1996).

1.7. Relationship between Statistics and ANN

- Major points of *differences* worth noting are:
  - While statistics is concerned with data analysis, supervised ANN emphasize statistical inference .
  - Some neural networks are not concerned with data analysis (e.g., those intended to model biological systems)
  - Some neural networks do not learn (e.g., Hopfield nets) and therefore have little to do with statistics.

- Some neural networks can learn successfully only from noise-free data (e.g., ART or the perceptron rule) and therefore would not be considered statistical methods

- Most neural networks that can learn to generalize effectively from noisy data are similar or identical to statistical methods.
- Major points of *similarities* worth noting are:
  - Feedforward nets with no hidden layer (including functional-link neural nets and higher-order neural nets) are basically generalized linear models.
  - Probabilistic neural nets are identical to kernel discriminant analysis.
  - Kohonen nets for adaptive vector quantization are very similar to k-means cluster analysis.
  - Hebbian learning is closely related to principal component analysis.

- Some neural network areas that appear to have no close relatives in the existing statistical literature are:
  - Kohonen's self-organizing maps.
  - Reinforcement learning (although this is treated in the operations research literature on Markov decision processes

---

2. Types of ANN

- There are many types of ANNs.
  - Many new ones are being developed (or at least variations of existing ones).

2.1. Networks based on supervised and unsupervised learning

2.1.1. Supervised Learning

- the network is supplied with a sequence of both input data and desired (target) output data network is thus told precisely by a "teacher" what should be emitted as output.
- The teacher can during the learning phase "tell" the network how well it performs ("reinforcement learning") or what the correct behavior would have been ("fully supervised learning").

2.1.2. Self-Organization or Unsupervised Learning

- a training scheme in which the network is given only input data, network finds out about some of the properties of the data set , learns to reflect these properties in its output. e.g. the network learns some compressed representation of the data. This type of learning presents a biologically more plausible model of learning.
- what exactly these properties are, that the network can learn to recognise, depends on the particular network model and learning method.

2.2. Networks based on Feedback and Feedforward connections

- The following shows some types in each category

- *Unsupervised Learning*
  Feedback Networks:
    - Binary Adaptive Resonance Theory (ART1)
    - Analog Adaptive Resonance Theory (ART2, ART2a)
    - Discrete Hopfield (DH)
    - Continuous Hopfield (CH)
    - Discrete Bidirectional Associative Memory (BAM)
    - Kohonen Self-organizing Map/Topology-preserving map (SOM/TPM)

  Feedforward-only Networks:
    - Learning Matrix (LM)
    - Sparse Distributed Associative Memory (SDM)
    - Fuzzy Associative Memory (FAM)
    - Counterprogation (CPN)
- *Supervised Learning*
  Feedback Networks:
    - Brain-State-in-a-Box (BSB)
    - Fuzzy Congitive Map (FCM)
    - Boltzmann Machine (BM)
    - Backpropagation through time (BPTT)

  Feedforward-only Networks:
    - Perceptron
    - Adaline, Madaline
    - Backpropagation (BP)
    - Artmap
    - Learning Vector Quantization (LVQ)
    - Probabilistic Neural Network (PNN)
    - General Regression Neural Network (GRNN)

---

3. Methodology: Training, Testing and Validation Datasets

- In the ANN methodology, the sample data is often subdivided into *training*, *validation,* and *test* sets.
- The distinctions among these subsets are crucial.
- Ripley (1996) defines the following (p.354):
  - *Training set:* A set of examples used for learning, that is to fit the parameters [weights] of the classifier.
  - *Validation set:* A set of examples used to tune the parameters of a classifier, for example to choose the number of hidden units in a neural network.
  - *Test set:* A set of examples used only to assess the performance [generalization] of a fully-specified classifier.

---

4. Application of a Supervised ANN for a Classification

Problem

- In this section, we describe how two neural networks to classify data and estimate unknown functions. Multi-Layer Perceptron (MLP) and fuzzy ARTMAP networks.

4.1. Multi-Layer Perceptron (MLP) Using Backpropagation

- A popular ANN classifier is the Multi-Layer Perceptron (MLP) architecture trained using the backpropagation algorithm.
- In overview, a MLP is composed of layers of processing units that are interconnected through weighted connections.
  - The first layer consists of the input vector
  - The last layer consists of the output vector representing the output class.
  - Intermediate layers called `hidden` layers receive the entire input pattern that is modified by the passage through the weighted connections. The hidden layer provides the internal representation of neural pathways.
- The network is trained using backpropagation with three major phases.
  - *First phase*: an input vector is presented to the network which leads via the forward pass to the activation of the network as a whole. This generates a difference (error) between the output of the network and the desired output.
  - *Second phase:* computethe error factor (signal) for the output unit and propagates this factor successively back through the network (error backward pass).
  - *Third phase:* compute the changes for the connection weights by feeding the summed squared errors from the output layer back through the hidden layers to the input layer.
- Continue this process until the connection weights in the network have been adjusted so that the network output has converged, to an acceptable level, with the desired output.
- Assign "unseen" or new data
  - The trained network is then given the new data and processing and flow of information through the activated network should lead to the assignment of the input data to the output class.
- For the basic equations relevant to the backpropagation model based on generalized delta rule, the training algorithm that was popularized by Rumelhart, Hinton, and Williams, see chapter 8 of Rumelhart and McClelland (1986).

4.1.1. Things to note while using the backpropagation algorithm

- *Learning rate:*
  - Standard backprop can be used for incremental (on-line) training (in which the weights are updated after processing each case) but it does not converge to a stationary point of the error surface. To obtain convergence, the learning rate must be slowly reduced. This methodology is called "stochastic approximation."
  - In standard backprop, too low a learning rate makes the network learn very slowly. Too high a learning rate makes the weights and error function diverge, so there is no learning at all.
  - Trying to train a NN using a constant learning rate is usually a tedious process

requiring much trial and error. There are many variations proposed to improve the standard backpropagation as well as other learning algorithms that do not suffer from these limitations. For example, stabilized Newton and Gauss-Newton algorithms, including various Levenberg-Marquardt and trust-region algorithms).

- *Output Representation:*
    - use 1-of-C coding or dummy variables.
    - For example, if the categories are Water, Forest and Urban, then the output data would look like this:

```
Category  Dummy variables
--------  ---------------
 Water       1   0   0
 Forest      0   1   0
 Urban       0   0   1
```

- *Input Data:*
    - Normalize or transform the data into [0,1] range. This can help for various reasons.
- *Number of Hidden Units:*
    - simply try many networks with different numbers of hidden units, estimate the generalization error for each one, and choose the network withthe minimum estimated generalization error.
- *Activation functions*
    - for the hidden units, are needed to introduce nonlinearity into the network.
    - Without nonlinearity, hidden units would not make nets more powerful than just plain perceptrons (which do not have any hidden units, just input and output units).
    - The sigmoidal functions such as logistic and tanh and the Gaussian function are the most common choices.

4.2. Fuzzy ARTMAP

- This is a supervised neural network architecture that is based on "Adaptive Resonance Theory", proposed by Stephen Grossberg in 1976.
- ART encompasses a wide variety of neural networks based explicitly on human information processing and neurophysiology.
- ART networks are defined algorithmically in terms of detailed differential equations intended as plausible models of biological neurons.
- In practice, ART networks are implemented using analytical solutions or approximations to these differential equations.
- ART is capable of developing stable clusterings of arbitrary sequences of input patterns by self-organisation.

- Fuzzy ARTMAP is based on ART.
    - Fuzzy ARTMAP's internal control mechanisms create stable recognition categories of optimal size by maximizing code compression while minimizing predictive error during on-line learning.
    - Fuzzy ARTMAP incorporates fuzzy logic in its ART modules
    - Fuzzy ARTMAP has fuzzy set-theoretic operations instead of binary set-theoretic

operations.
  - It learns to classify inputs by a fuzzy set of features (or a pattern of fuzzy membership values between 0 and 1)

4.2.1. Basic architecture of fuzzy ARTMAP

- A pair of fuzzy ART modules, ART_a and ART_b, connected by an associative learning network called a *map field*
  - the map field makes the association between ART_a and ART_b categories.
- A mismatch between the actual and predicted value of output causes a memory search in ART_a, a mechanism called *match tracking*
- *Vigilance*, a parameter (0-1) in ART_a, is raised by the minimum amount necessary to trigger a memory search.
  - This can lead to a selection of a new ART_a category that is a better predictor of output.
- Fast learning and match tracking enable fuzzy ARTMAP to learn to predict novel events while maximizing code compression and preserving code stability.
- Carpenter (1997) gives a complete description of the algorithm and description of fuzzy ARTMAP for remote sensing applications.

4.3. Software

- There are many commercial and free software packages for running backpropagation.

- Fuzzy ARTMAP
  - The exercise at the end of this chapter uses a classification example using this package. It uses data from remote sensing data set.

---

5. Application Exercises: Backpropagation algorithm X Fuzzy ARTMAP for Classification of Landcover Classes

5.1. Data Set 1

- NOTE: please contact the author to obtain a copy of this data
- This data set has 6 inputs (Landsat TM Spectral Bands) and 8 output classes represented as a single number (1-8) for each pixel.
- Train the neural network with 80% of the data and test it on the remaining 20% of the data.
  1. Compare the performance of backpropagation and fuzzy ARTMAP.
  2. Use different settings of crucial parameters such as learning (in backpropagation) and vigilance (In fuzzy ARTMAP). Are results different?
  3. Use a conventional statistical models to benchmark the performance of neural networks?

5.2. Data Set 2

- This data is found in http://lib.stat.cmu.edu/datasets/boston.
    1. Use MLP with backpropagation to estimate the Median value of owner-occupied homes in Boston.
    2. Use a conventional regression model to compare your results?

---

6. Summary

- This unit has introduced some definitions and types of neural networks
    - It has examined the differences between ANN and statistics
    - It has given an overview of application domains
    - It has provided the use of MLP and Fuzzy ARTMAP neural networks for classification problems
    - Sample data sets are provided along with information on free software sources to enable users to learn the applications of ANN.

---

7. Review and Study Questions

---

8. References

8.1. References in the text of this unit

- Fischer, M.M. Expert systems and artificial neural networks for spatial analysis and modeling. Essential components for knowledge-based geographic information systems, Paper presented at the Specialist Meeting on `GIS and Spatial Analysis' organized by the NCGIA, San Diego, April 15-18, 1992.

- Fischer, M., and Gopal, S. Neural network models and interregional telephone traffic: comparative performances between multilayer feedforward networks and the conventional spatial interaction model, *Journal of Regional Science*, 34,4, 503-527, 1995.

- Carpenter, G., Gjaja, M., Gopal, S., and Woodcock, C. ART networks in Remote Sensing, *IEEE Transactions on Geoscience and Remote Sensing*, 35(2), 308-325, 1997.

- Gopal, S. and Fischer, M. Learning in single hidden layer feedforward neural network models: backpropagation in a spatial interaction modeling context, Geographical

Analysis, 28 (1), 38-55, 1996.

- Gopal, S. and Woodcock, C. E. Remote sensing of forest change using artificial Neural Networks, *IEEE Transactions on Geoscience and Remote Sensing*, 34 (2), 398-404, 1996.
- Hopfield, J.J.and Tank, D.W. (1985): Neural computation of decisions in optimization problems, *Biological Cybernetics* 52, pp. 141-152.
- Openshaw, S. (1993). Modelling spatial interaction using a neural net, in M. M. Fischer and P. Nijkamp (eds) *GIS Spatial Modeling and Policy*, Springer, Berlin, pp. 147-164.

8.2. Books

- For the interested reader, I have selected some books out of a plethora of publications. This list is not exhaustive but a good starting point.

- Bishop, C.M. (1995). Neural Networks for Pattern Recognition, Oxford: Oxford University Press. ISBN 0-19-853849-9 (hardback) or 0-19-853864-2 (paperback), xvii+482 pages.
- Hertz, J., Krogh, A., and Palmer, R. (1991). Introduction to the Theory of Neural Computation. Addison-Wesley: Redwood City, California. ISBN 0-201-50395-6 (hardbound) and 0-201-51560-1 (paperbound)
- Ripley, B.D. (1996) Pattern Recognition and Neural Networks, Cambridge: Cambridge University Press, ISBN 0-521-46086-7 (hardback), xii+403 pages.
- Weigend, A.S. and Gershenfeld, N.A., eds. (1994) Time Series Prediction: Forecasting the Future and Understanding the Past, Addison-Wesley:Reading, MA.
- Masters, Timothy (1994). Practical Neural Network Recipes in C++, Academic Press, ISBN 0-12-479040-2, US $45 incl. disks.
- Fausett, L. (1994), Fundamentals of Neural Networks: Architectures, Algorithms, and Applications, Englewood Cliffs, NJ: Prentice Hall, ISBN 0-13-334186-0. Also published as a Prentice Hall International Edition, ISBN 0-13-042250-9. Sample software (source code listings in C and Fortran) is included in an Instructor's Manual.
- Michie, D., Spiegelhalter, D.J. and Taylor, C.C. (1994), Machine Learning, Neural and Statistical Classification, Ellis Horwood.
- Aleksander, I. and Morton, H. (1990). An Introduction to Neural Computing. Chapman and Hall. (ISBN 0-412-37780-2).

8.3. Classics

- Kohonen, T. (1984). Self-organization and Associative Memory. Springer-Verlag: New York. (2nd Edition: 1988; 3rd edition: 1989).

- Rumelhart, D. E. and McClelland, J. L. (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition (volumes 1 & 2). The MIT Press.

Citation

To reference this material use the appropriate variation of the following format:

> Sucharita Gopal. (1998) Artificial Neural Networks for Spatial Data Analysis, *NCGIA Core Curriculum in GIScience*, http://www.ncgia.ucsb.edu/giscc/units/u188/u188.html, posted December 22, 1998.

Created: November 23, 1998. Last revised: December 22, 1998.

Unit 188 - Artificial Neural Networks for Spatial Data Analysis

Table of Contents

Unit 188 - Artificial Neural Networks for Spatial Data Analysis

Metadata and Revision History

Page contents

---

1. About the main contributor

- author: Sucharita Gopal, Department of Geography and Centre for Remote Sensing, Boston University, Boston MA 02215

2. Details about the file

- unit title: Artificial Neural Networks for Spatial Data Analysis
- unit key number: Unit 188

3. Key words

4. Index words

5. Prerequisite units

6. Subsequent units

7. Revision history

- November 23, 1998 - original draft created
- December 22, 1998 - original draft posted
