2017-04-05

Artificial neural networks in geospatial analysis

Sucharita Gopal. 2017. "Artificial Neural Networks in Geospatial Analysis." In: The International Encyclopedia of Geography: People, the Earth, Environment, and Technology. Edited by Douglas Richardson, Noel Castree, Michael F. Goodchild, Audrey Kobayashi, Weidong Liu, and Richard A. Marston. DOI: 10.1002/9781118786352.wbieg0322

https://hdl.handle.net/2144/29237

Artificial neural networks in geospatial analysis

Sucharita Gopal
Boston University, USA

Artificial neural networks are computational models widely used in geospatial analysis for data classification, change detection, clustering, function approximation, and forecasting or prediction. There are many types of neural networks based on learning paradigm and network architectures. Their use is expected to grow with increasing availability of massive data from remote sensing and mobile platforms.

Key Words geospatial; neural networks; remote sensing; space-time; spatial analysis; spatial database; spatial modeling; spatiotemporal data; supervised learning; technology; unsupervised learning.

Background

Artificial neural networks (ANN) are computational models inspired by and designed to simulate biological nervous systems that are capable of performing specific information processing tasks such as data classification and pattern recognition. ANN seeks to replicate the massively parallel nature of a biological neural network. A neural network is a system composed of many simple processing nodes whose function is determined by network structure and connection strengths. Similar to biological neural networks, AAN can acquire knowledge through a learning process. Interneuron connection strengths known as synaptic weights are used to store the knowledge and make it available for use. Each artificial neuron receives one or more input signals and sums them to generate an output. Usually the sums of each node are weighted, and the values are passed through a nonlinear function known as an activation function or transfer function. If a weighted sum exceeds some predetermined threshold value, then an excitatory output (1) is produced. Otherwise, an inhibitory output (−1) is produced. The transfer function can be a sigmoid function or other nonlinear, piecewise linear or step function. Researchers in a variety of disciplines including engineering, psychology, mathematics, and physics have contributed to the explosive growth in the field that has continued to this day. Two standard texts in the field are Bishop (1995) and Ripley (1996).

Neural networks can be used to address a wide variety of real-world problems. They have the ability to learn from experience in order to improve their performance and dynamically adapt themselves to changes in the environment. In addition, they are able to deal with fuzzy or incomplete information and noisy data, and can be very effective, especially in situations where it is not possible to define the rules or steps that lead to the solution of a problem. Hence they are fault tolerant. In addition, the ANN information processing model is inherently parallel. Today ANNs are used in a variety of disciplines including engineering, finance, artificial perception, and control and simulation.

The wide use of ANN in geospatial science stems from their roles in spatial data processing and analysis. Satellites orbiting and imaging the earth produce massive amounts of geospatial data, on the order of tera- to peta-bytes. ANN, programmed on parallel neurally inspired hardware architectures, can analyze and classify this vast amount of data quickly and draw meaningful insights via mapping or modeling. ANN’s generalization capability in dealing with classification across multiple spatial scales and resolutions is a significant advantage. ANN can also deal with fuzzy data as well as qualitative spatial data. Moreover, many physical processes modeled in the geospatial sciences require accurate knowledge of process dynamics and such knowledge is often unknown. In this case, ANN can be used in function approximation.

The learning that occurs in ANN is not affected by the integration of multisource data. The learning process is robust and fault tolerant. Newer learning paradigms such as *semi-supervised learning* or *self supervised learning* can deal with incomplete data, overcoming the difficulties and expenses involved in gathering training labels for geospatial sensor data. Recent studies have attempted to make ANN more spatially explicit, either by introducing fundamental spatial principles such as spatial autocorrelation directly into the neural network structure or during post-processing and labeling using spatial neighborhood relationships. When using an ensemble approach, ANN can assist in characterizing the spatial heterogeneity of the Earth's surface and spatial uncertainty in labeling.

Types of ANN

There are several types of ANN based on the learning paradigm, architecture, and function.

Learning paradigm – supervised, unsupervised, and semi-supervised learning

Supervised learning: In supervised learning, the ANN is supplied with a sequence of both input data and desired (target) output data; the network is thus told precisely by a "teacher" what should be emitted as output. During the learning phase, the teacher can "instruct" the network about how well it performs ("reinforcement learning") or what the correct behavior would have been ("fully supervised learning"). The ANN learns the association between input and output classes until some criterion of successful learning is met. A commonly used metric is the mean squared error, which tries to minimize the average squared error between the network's output and the target value over all the input-output training data pairs. The well-known backpropagation algorithm for training tries to minimize this cost using a gradient descent function.

In the ANN methodology, the sample data for supervised training is often subdivided into *training*, *validation*, and *testing* sets. The distinctions among these subsets are crucial. Ripley (1996, 354) defines the following: Training data is a set of examples used for learning that is to fit the parameters [weights] of the classifier. Validation data is 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, while testing data is a set of examples used only to assess the performance [generalization] of a fully specified classifier.

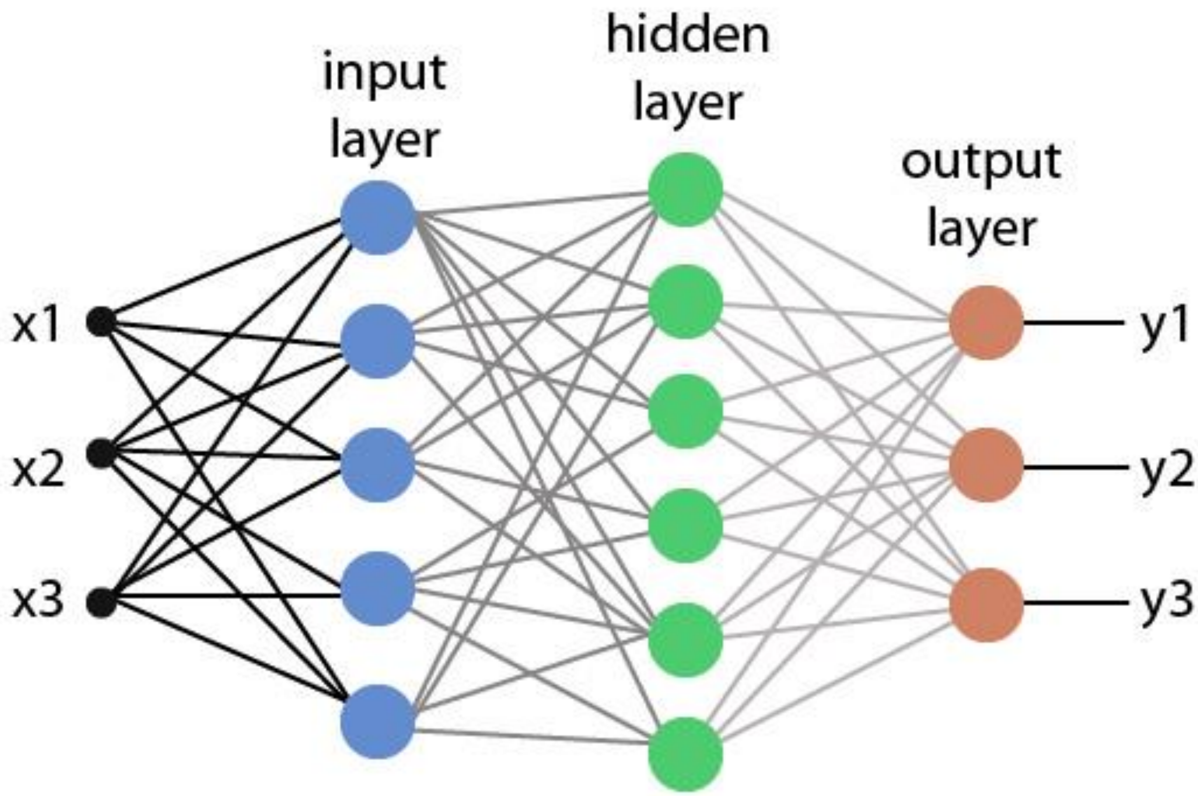
In unsupervised learning, the training scheme only consists of input data. The ANN discovers some of the properties of the dataset and learns to reflect these properties in its output. The issue of what exactly these properties are, that the network can learn to recognize, depends on the particular network model and learning method. Unsupervised ANNs can discover the underlying structure of the data, as well as encode, compress, and transform data values. This type of learning presents a biologically more plausible model of learning. The self-organizing map (SOM), a well-known unsupervised ANN, learns to produce a low-dimensional discretized representation of the input space using a neighborhood function to preserve its topological properties.

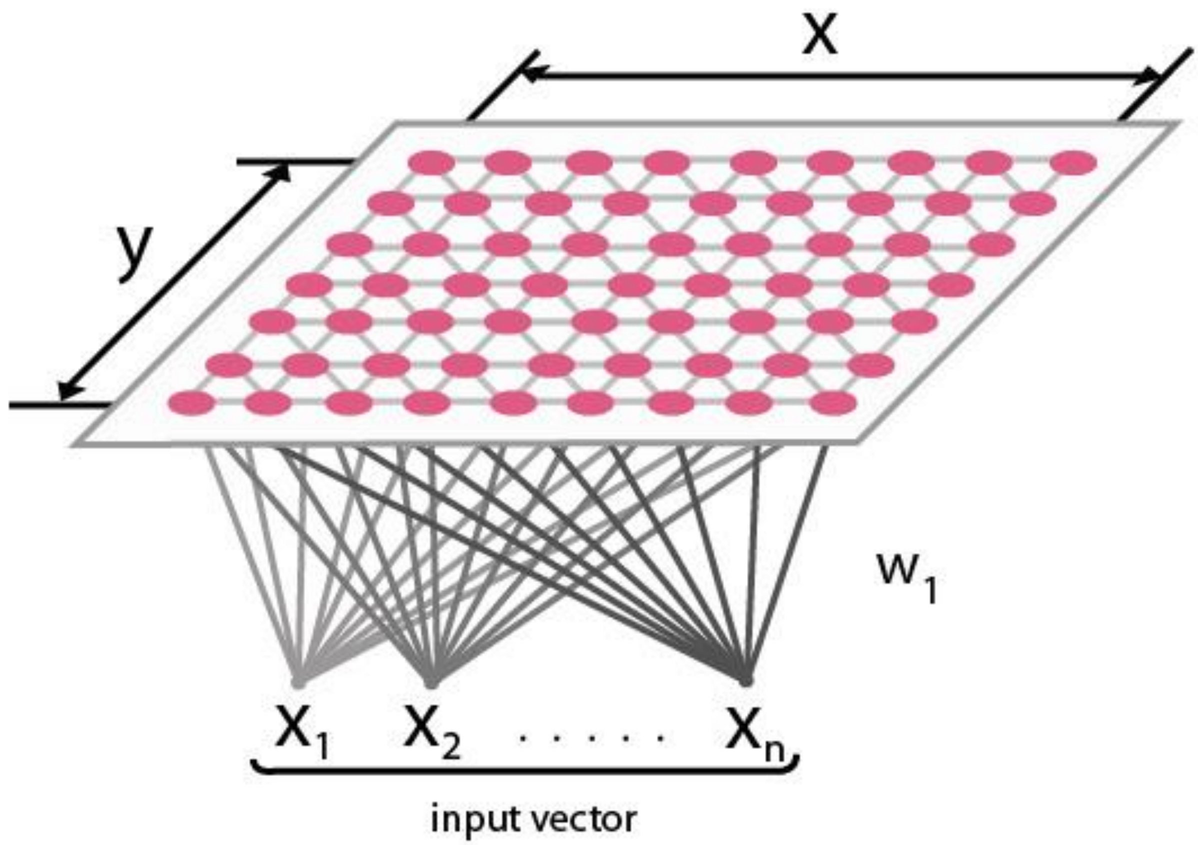
Real-world applications of supervised ANN are popular but face a number of practical challenges, such as scalability and limited availability of labeled training data. In spatial mapping, obtaining labeled datasets is expensive, and sometimes impossible. A recent solution is the development of a new learning paradigm called *semi-supervised learning (SSL).* In SSL, the user strategically selects the data to be manually labeled, and then lets the ANN iteratively retrain itself on its own output using the remaining unlabeled data. A variation of SSL is the Self-Supervised ARTMAP, which uses only a subset of input features during training. This is a great advantage in GIS and remote sensing, as new information about Earth's features is constantly being acquired.

ANN network architecture

Figure 1a shows a typical structure of an ANN network consisting of processing units (neurons) arranged in two layers, and links (synapses) between the processing units in the different layers. There are input nodes and output nodes representing inputs and outputs in a supervised ANN. In addition, there are hidden units that can vary depending on the nature of the problem. Experimentation is often required to determine the best number of hidden units. Too many hidden units may lead the network to overfit the training data, thus reducing generalization accuracy. On the other hand, too few hidden units may prevent the network from being able to learn the required function. Each connection between nodes has a weight associated with it that is adjusted during learning. In addition, an activation function converts the processing unit's weighted input to its output activation.

<FIG 1a and 1b>





ANN network function – feedforward and recurrent networks
ANNs are also differentiated based on whether there are cycles or loops in the network. In the feedforward network (Figure 1a), information signals move in one direction, from the input nodes, through the hidden nodes (if any), to the output nodes. There are no cycles or loops in the network. Hence these are called *acyclic graphs*. An example is a multilayer feedforward neural network. In contrast, recurrent networks (Figure 1b) are cyclic graphs since they contain cycles or loops. Some recurrent networks (e.g., ART) can be extremely complicated.

Three ANN models

Multilayer perceptron using backpropagation

A popular ANN classifier is the feedforward multilayer perceptron (MLP) architecture. An MLP is composed of layers of processing units in a directed graph that are linked through weighted connections (Figure 1a). The first and last layers consist of the input variables (e.g., spectral bands) and output classes. The intermediate layers, called hidden layers, provide the internal representation of neural pathways. An MLP ANN maps sets of input data onto a set of appropriate outputs. Except for the input nodes, each processing unit has a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation to train the network.

Learning occurs in the MLP by modifying connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result. This process is continued until the weights in the network have been adjusted such that the network output has converged, to an acceptable level, with the desired output. In the next phase, the fully trained network is given new data, and the processing and flow of information through the now-activated network should lead to the assignment of the input data to an output class.

Self-organizing map

The self-organizing map (SOM) is a popular tool for mapping and clustering high dimensional data (Figure 1b). In a SOM, neurons or units are arranged in a rectangular or hexagonal group of units of predefined dimensions ($m \times n$ rows and columns). The number of units is usually small relative to the dimensionality of the input data. The SOM algorithm computes an ordered mapping, a kind of projection that forces each of the input data records to map onto a defined grid during the iterative learning process. The goal of the SOM is to preserve the similarities between samples such that similar input data records map on to the same or neighboring units in the grid, while dissimilar data records are mapped on to non-neighboring units. Thus SOM can incorporate spatial neighborhood and spatial autocorrelation effects that are commonly encountered in GIS and spatial analysis. Over the last decade, SOM has been increasingly used in a supervised fashion.

At the start of training, the grid units are initialized via a random sample of p observations from the input data. Each grid unit is characterized by a codebook vector that describes the typical pattern of that unit. The aim of the SOM algorithm is to update codebook vectors so that the input data are best described by the small number of grid units. During training, the distance between phenomena in the presented dataset and the codebook vectors for each of the units is determined by a distance measure. The unit whose codebook vector is closest or most similar to the presented sample is the winning unit or node. The winning nodes' codebook vector is then updated so that the winning unit is made more similar to the presented sample.

Self-Supervised ARTMAP

Self-Supervised ARTMAP is based on Fuzzy ARTMAP (Carpenter 2013). Fuzzy ARTMAP is based on "Adaptive Resonance Theory" (ART), proposed by Stephen Grossberg in 1976. Fuzzy ARTMAP's internal control mechanisms create stable recognition categories of optimal size by maximizing code compression while minimizing predictive error during online learning. Fuzzy ARTMAP incorporates fuzzy logic in its ART modules. ART overcomes the stability-plasticity dilemma; namely, how a neural network can learn quickly about new objects and events (plasticity) without overwriting or forgetting previously learned memories (stability).

In the first phase of self-supervised learning, the ARTMAP system is trained on labeled data with a limited set of labeled data (input features and output classes), similar to learning in a traditional supervised ANN. In the next phase, the system continues to learn using an expanded set of input features with no labels or output classes. This "unsupervised learning" enables the system to build on its existing knowledge with new information, resulting in improved accuracy, compared to that of the initial trained system, without worsening its performance. This is demonstrated by Carpenter (2013) on a remote sensing (Landsat) database of land cover related to the greater Boston region.

Review of ANN applications in geospatial science

ANN models have been used in the geospatial sciences since the 1990s. A literature review was conducted using Google Scholar to identify existing publications of relevance with key terms "*neural networks and GIS*," "*spatial analysis,*" "*spatial interaction,*" or "*neural networks and remote sensing*." The search was done with a five-year interval starting in 1990. A majority of publications were peerreviewed journal articles, conference reports, and technical papers. Four key findings are: first, the number of papers using neural networks in remote sensing was much larger, likely because remote sensing is multidisciplinary, including fields such as engineering, mathematics, and physics. The number of papers in GIS and spatial analysis was smaller, perhaps because the broad search terms do not include applications in the transportation and energy sectors. Second, there has been a steady increase in the numbers of papers in both fields, indicating a broader dissemination of neural network modeling in the last two decades. Third, an analysis of broad themes in both fields over the last three decades indicates that a vast majority of publications (around 60%) relate to *classification* and *clustering*, followed by *function approximation*, *estimation*, and *optimization*. The last category, *optimization*, is less prevalent in the remote sensing field. Fourth, *function approximation* and *estimation* applications are popular in climatology and hydrology and in modeling physical processes.

Common applications of ANN

Let us consider the most common applications of ANN geospatial science – *classification*, *change detection*, *clustering*, *function approximation*, and *forecasting* or *prediction*.

The goal of classification is to assign an input pattern (like spectral bands or sociodemographics) represented by an input feature vector to one of several output classes. The best known application is land use/land cover classification in remote sensing and GIS involving multisensor, multidate, multisource, and multiscale data. The number of inputs can vary up to hundreds, given the data capture capabilities of NASA’s Earth Observing System Data and Information System (EOSDIS). Classification is typically performed using supervised learning methods. In many studies, the performance of ANN has been shown to be superior to that of traditional classifiers such as Maximum Likelihood (Carpenter *et al.* 1997).

The measurement and characterization of changes happening on the Earth’s surface at a variety of spatial and temporal scales has long been of interest to geospatial scientists. In this context, ANN has helped to identify patterns of change through time as well as to distinguish abrupt (deforestation or forest fires) and continuous (coastal erosion) changes. This application involves mostly supervised neural networks or hybrid approaches, where classification is done at two time periods and the differences are then classified and mapped.

The main objective of clustering is to group large amounts of data into meaningful categories that can be disjoint, overlapping, or organized in some hierarchical fashion. A category or cluster is characterized by maximum similarity among its members and minimum similarity between its members and the ones belonging to other classes. Unsupervised ANNs are often used in this context, focusing on exploratory data analysis and visualization of clusters extracted from massive data with many dimensions.

Suppose an unknown function $\Phi(x)$ (subject to noise) has generated a set of **n** input-output pairs $\{(x_1,y_1), (x_2,y_2), .., (x_n,y_n)\}$. The task of function approximation is to find an estimate □ Φ^* of the unknown function Φ □^. ANN can approximate functions with arbitrarily high degrees of nonlinearity with a sufficient number of nodes and layers. These advantages have led to the recent use of ANN estimation algorithms for geophysical parameter retrievals. Typically, supervised ANNs are used in this context to solve forward and inverse remote sensing problems.

The goal of optimization is to find a solution that satisfies a set of constraints such that an objective function is maximized or minimized. The most famous problem relevant in GIS is the *Traveling Salesman Problem*, an *NP*-complete problem where a traveling salesman must visit **N** number of cities with the following constraints: start and end the trip at the same location, visit each city only once during the trip, and minimize distance traveled during a trip. The order of visits in the trip does not matter for finding the solution. The ANN used in this context is the Hopfield model. Related modern applications include vehicle routing (such as FedEx, UPS) where the goal is to minimize the number of vehicles involved in delivery, total travel time, and total delivery time.

Given a set of **n** samples {(y(t_1), (y(t_2), .., (y(t)$_n$)} in a time series, t_1, t_2, …, t_n, the task is to predict the value y(t_{n+1}) at some future time t_{n+1}. Two important issues must be addressed in this application: the frequency with which data should be sampled, and the number of data points which should be used in the input representation. These issues are settled empirically in most application contexts, but results from work in complex dynamic systems have employed heuristics. Typical examples of this supervised ANN approach are market predictions (price of crude oil or gold), climatological (modeling ENSO events using large-scale climatological parameters), and network traffic forecasting (spatial interaction in terms of flows between origin and destination pairs).

Present technology trends and the future of ANN

During the last decade, there have been changes in both ANN hardware and software. The new processors used in real-time ANN consist of electronic components that can be connected by circuits that are designed to mimic biological synapses. They are known as “neuromorphic” processors since they are based on large groups of neuron-like elements. The development of “deep learning” algorithms that can be embedded in these neuromorphic chips has led to substantial growth in real-time applications. These developments will vastly impact geospatial sciences in the next decade.

References

Bishop, C.M. 1995. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.

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

Carpenter, G. 2013. “ART, Cognitive Science, and Technology Transfer.” Wiley Interdisciplinary Reviews (WIREs). *Cognitive Science*, 4(6): 707–719. DOI.org/10.1002/wcs.1260.

Ripley, B.D. 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.