SOFT COMPUTING-NEURAL NETWORKS ENSEMBLES

[1]K.Venu Gopala Rao, [2]P.Prem Chand, [3]M.V.Ramana Murthy
[1]Department of Computer Science&Engg, GNITS, Hyderabad-500 008, A.P.
[2]Department of Computer Science&Engg, Univ.College of Engineering, Osmania University, Hyderabad- 500 007, A.P.
[3]Department of Computer Science&Engg, Univ.College of Engineering, Osmania University, Hyderabad- 500 007, A.P.

ABSTRACT

Neural Network ensemble is a learning paradigm where a collection of finite number of neural networks is trained for the same task. It is understood that the generalization ability of neural networks, i.e., training many neural networks and then combining their predictions. ANN ensemble techniques have become very popular amongst neural network practitioners in a variety of ANN application domains. There are many different ensemble techniques, but the most popular include some elaboration of *bagging and boosting or stacking.* When applied to Ann's, ensemble techniques can produce dramatic improvements in generalization performance.

Since this technology behaves remarkably well, recently it has become a very hot topic in both neural networks and machine learning communities, and has already been applied to diversified areas such as face recognition, optical character recognition, etc. In general, a neural networks ensemble is constructed in two steps, i.e., training a number of component neural networks, then combining the component predictions.

Key words: ANN*, learning paradigm, ensemble, bagging, boosting, component predictions*

1. INTRODUCTION

Soft Computing is an emerging field that consists of complementary elements of fuzzy logic, neural computing, evolutionary computation, machine learning and probabilistic reasoning. Due to their strong learning, cognitive ability and good tolerance of uncertainty and imprecision, soft computing techniques have found wide applications.

Generally speaking soft computing techniques resemble human reasoning more closely than traditional techniques which are largely based on conventional logical systems, such as sentential logic and predicate logic, or rely heavily on the mathematical capabilities of a computer.

Unlike hard computing schemes that strive for exactness and for full truth, soft computing techniques are tolerant of imprecision, partial truth, and uncertainty. They are mainly, similar to those methods that are applied by human mind in order to reach decisions, and conclusions. The guiding principle of soft computing is: Exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness and low solution cost.

Recent advances in supervised image classification have shown that conventional hard classification techniques, which allocate each pixel to a specific class, are often inappropriate for applications where mixed pixels are abundant in the image. Typical applications where the classes may not be considered to be discrete and mutually exclusive, is land cover mapping form remotely sensed data. For instance, in a remotely sensed image, there may be continuous and gradual transition between land cover classes resulting in regions of mixed pixel composition. For this reason fuzziness should be accommodated in classical procedure. Recent progress in neural techniques has demonstrated the usefulness of neural networks in variety of areas including classification of remote sensed images.

The theory of multi criteria decision-making has witnessed the birth of many new paradigms in last few years. The common feature of all the paradigms in multi criteria decision-making is that we need somewhere a fundamental operation called aggregation. Most common aggregation tools used today are weighted arithmetic means. However, some thirty years back, Sugeno proposed a concept of fuzzy integral. The distinguishing feature of fuzzy integral is that it can represent certain kind of interaction between criteria. However, the richness of fuzzy integral has to be paid by complexity of the model since number of coefficients grows exponentially with number of criteria to be aggregated. Here, we look for an efficient way for improving accuracy of classification using fuzzy integration and also with less computational complexity.

This study will particularly discuss information fusion in the sense of combining the results from multiple classifiers. Many researchers have earlier applied the fuzzy integral for image classification. The fuzzy integral operates on fuzzy measures, which stand for importance of each classifier of the degree of importance of collection of classifiers. It is observed that assigning a fixed valued for each fuzzy density is less effective due to the fact that each classifier may not perform equally well in recognizing all classes i.e., one particular classifier may recognize well certain type of classes but fail to recognize other classes. Therefore the weight of importance of fuzzy density of each classifier will need to be adjusted based on input type and simultaneous information given by other classifiers. The detailed discussion regarding the computation of fuzzy densities is covered in subsequent chapters. The rest of this chapter gives introduction about remote sensing and image classification, objective of this project and organization of rest of the report.

2. REMOTE SENSING

Remote sensing is a technique used to collect data about the earth without taking a physical sample of the earth's surface. A sensor is used to measure the energy reflected from the earth. This information can be displayed as a digital image or as a photograph. Sensors can be mounted on a satellite orbiting the earth, or on a plane or other airborne structure.

In much of remote sensing, the process involves an interaction between incident radiation and the targets of interest. This is exemplified by the use of imaging systems where the following seven elements are involved.

Energy Source or Illumination - the first requirement for remote sensing is to have an energy source, which illuminates or provides electromagnetic energy to the target of interest.

Radiation and the Atmosphere - as the energy travels from its source to the target, it will come in contact with and interact with the atmosphere it passes through. This interaction may take place a second time as the energy travels from the target to the sensor.

Interaction with the Target - once the energy makes its way to the target through the atmosphere, it interacts with the target depending on the properties of both the target and the radiation.

Recording of Energy by the Sensor - after the energy has been scattered by, or emitted from the target, we require a sensor (remote - not in contact with the target) to collect and record the electromagnetic radiation.

Transmission, Reception, and Processing - the energy recorded by the sensor has to be transmitted, often in electronic form, to a receiving and processing station where the data are processed into an image (hardcopy and/or digital).

Interpretation and Analysis - the processed image is interpreted, visually and/or digitally or electronically, to extract information about the target, which was illuminated.

Application - the final element of the remote sensing process is achieved when we apply the information we have been able to extract from the imagery about the target in order to better understand it, reveal some new information, or assist in solving a particular problem.

There are two basic types of sensors: passive and active sensors. Passive sensors record radiation reflected from the earth's surface. The source of this radiation must come from outside the sensor. In most cases, this is solar energy. Because of this energy requirement, passive solar sensors can only capture data during daylight hours. Active sensors are different from passive sensors. Unlike passive sensors, active sensors have their own source of energy to illuminate the target. For example, a laser-beam remote sensing system is an active sensor that sends out a beam of light with a known wavelength and frequency. This beam of light hits the earth and is reflected back to the sensor, which records the time it took for the beam of light to return. The images obtained from satellites are generally in multiple wavelengths.

2.1 Applications of Remote sensing

Remote Sensing data both in large form and digital format is utilized for deriving information about resources either adapting visual interpretation or computed aided analysis. Both types of measure require certain amount of ground support information. This information is normally termed to as "GROUND TRUTH". Using the ground truth the remote sensing data are analyzed interpreted and maps related to resources are generated. In the following graphs application potentials of remote sensing techniques related to various disciplines are described.

Geology and geographic mapping-Preparation of small scale maps of unmapped inaccessible areas, updating the existing geological maps, rapid preparation of tectonic maps, identification of feature favorable for mineral localization and ecological studies related to volcanoes glaciers over a long periods.

-Soil mapping
-Water resources
-Agriculture (crops)
-Forestry and vegetation mapping
-Land use mapping
-Other applications

Archaeology
Anthropology
Weather and climate
Cultural Resource management
Marine Engineering
Water Resource Management
Engineering Applications
Rangeland Applications
Civil Applications

3. IMAGE CLASSIFICATION

Image classification is the process of creating thematic maps from satellite imagery. The overall objective of the image classification is to automatically categorize all pixels in an image into land cover classes or themes. A thematic map is an informational representation of an image, which shows the spatial distribution of a particular theme. Themes can be diversified as their areas of interest. Examples of themes are soil, vegetation, water, and atmosphere.

A human analyst trying to classify features in an image uses the visual interpretation to identify homogenous group of pixels, which represent various features of interest (themes). Digital image classification uses the spectral information represented by digital numbers in one or more spectral bands. This type of image classification is called spectral image classification. Spectral classes are groups of pixels, which are similar (or near similar) in their brightness values in various spectral bands. The major challenge is to map the spectral classes to information classes. Information classes are those categories of interest that the analyst is actually trying to identify in the imagery, such as different kinds of crops, soils, etc. Basically image classification algorithm besides the spectral information also uses statistical information, which describes the features of image such as texture. Also the shape of the cluster is used in mapping the spectral classes to information classes. Most of the image classification algorithms generally use either statistical methods, Neural networks, Fuzzy logic, or Neuro-fuzzy approaches.

3.1 Classification Techniques

Broadly image classification algorithms can be classified as:

1 Supervised image classification
2 Unsupervised image classification

In supervised image classification, the human analyst identifies samples of different themes of interest. These samples are referred to as training data. Here the analyst is supervising the classification of image into themes, hence the name supervised classification. Based on these training samples the algorithm finds the statistical vector such as mean vector and spectral signature associated with each class. Whenever a new pixel is given to algorithm, it classifies the pixel to the class to which it closely resembles in terms of signature. In unsupervised classification the process is reversed, the pixels are first grouped solely based on similarity of associated spectral and statistical features.

In an unsupervised classification, the identities of land cover types to be specified, as classes within a scene are not generally known apriori because ground truth is lacking or surface features within the scene are not well defined. This classification is performed using clustering methods. Clustering algorithms are used for grouping the data based on the numerical wants pixels to be grouped into.

3.2 Supervised classification techniques

Supervised classification is the procedure most often used for interpreting the remote sensing image data. It rests upon using suitable algorithms to label the pixels in an image as representing particular ground cover types or classes.

The image analyst supervises the pixel categorization process by specifying to the computer algorithm, numerical descriptors of the various land cover types present in a scene. To do this, representative sample sites of known cover type called training areas are used to compile a numerical interpretation key that describes the spectral attributes for each feature type of interest. Each pixel in the data set is then compared numerically to each category in the interpretation key and labeled with the name of the category it looks most like.

The essential steps for supervised classification are:

1. Decide the set of ground cover types (such as water, urban regions, crop lands, range lands etc.,) into which the image is to be segmented.
2. Choose representative or prototype pixels from each of the desired set of classes. These pixels are said to form training data. The analyst identifies representative training areas and develops a numerical description of the spectral description of the spectral attributes of each land cover type of interest in the scene. Training sets of each class can be established using site visits, maps, air photographs or even photo interpretation of a color composite product formed from the image data. The training pixels for a given class will lie in a common region enclosed in a border. That region is called a training field.
3. Use the training data, signature of each class i.e., the parameters such as mean, variance, covariance matrix and correlation matrix are established. Next is the classification stage.
4. Using the trained classifier, label or classify every pixel in the image into one or the cosmired ground cover types. Each pixel in the image data set is categorized into land cover class is most closely resembles. If the pixel is sufficiently similar to any training data set is usually labeled unknown. The category label assigned to each pixel in this process is then record in the corresponding cell of the interpreted data set (an output image.). Thus the multidimensional image matrix is used to develop a corresponding matrix of interpreted land cover category types. After entire data set has been categorized, the results are presented in the output stage.
5. Being digital in character, the results may be used in a number of different ways. These typical forms of output products are thematic maps, tables of full scene or sub scene area statistics for various land cover classes, and digital data files amenable to inclusion in a Geographic Information System (GIS). In this latter case, the classification output becomes a GIS input.

Various supervised classification methods may be used to assign an unknown pixel to one of a number of classes. The choice of a particular classifier or decision rule depends on the nature of the input data and the desired output. Among the most frequently used classification algorithms are the maximum likelihood, Bayesian, minimum distance and parallelepiped algorithms.

4. NEURAL NETWORK ENSEMBLES

Neural Network ensemble is a learning paradigm where a collection of finite number of neural networks is trained for the same task. It originates from Hansen and Salamon's work, which shows that the generalization ability of neural networks, i.e., training many neural networks and then combining their predictions.

ANN ensemble techniques have become very popular amongst neural network practitioners in a variety of ANN application domains. There are many different ensemble techniques, but the most popular include some elaboration of *bagging and boosting or stacking* [23]. When applied to ANNs, ensemble techniques can produce dramatic improvements in generalization performance. The underlying idea of all these techniques is to generate multiple versions of a predictor, which when combined, will provide "smoother" more stable predictions.

Since this technology behaves remarkably well, recently it has become a very hot topic in both neural networks and machine learning communities, and has already been applied to diversified areas such as face recognition, optical character recognition, etc. In general, a neural networks ensemble is constructed in two steps, i.e., training a number of component neural networks, then combining the component predictions.

Bagging (an abbreviation of "bootstrap aggregation") is one of the most popular ANN ensemble techniques. It uses the bootstrap, a very popular statistical re sampling technique, to generate multiple training sets and networks for an ensemble (see Figure 4 for an illustration of this). Although other ensemble techniques such as boosting have been shown to out-perform bagging on some data-sets, bagging has a number of key advantages when applied to real-world tasks decision support. One of the most important is the ease with which confidence intervals can be computed. Another is the robustness and stability of the technique itself—Breiman showed that it will always perform at least as well as an individual predictor, as long as the predictor is unstable .

Methods for Combining set of Neural Networks

The combination strategy for a group pf neural network is of fundamental importance, and can decide the performance of the whole system. Two broad types of combination strategies are

1 Linear Combinations
2 Non linear Combinations

Linear Combinations

The simplest possible is to take a linearly weighted summation of individual predictor outputs. The output of combination is :

$$f_{res} = \Sigma^{M}_{i=1} w_i f_i \quad (4.1)$$

where M is number of neural networks, f_i is the output of i th neural networks, and w_i is corresponding non-negative real values combination weight. This is also referred to as linear fusion strategy. Specializations of the linear combinations are the convex combination, where weights can be non-uniform but have constraint that they sum to one: $\Sigma^{M}_{i=1} w_i = 1.0$; and also the result of combination in 3.2 where all weights are equal at

$w_i = 1/M$

$$f_{res} = (1/M)\, \Sigma^{M}_{i=1} f_i \quad (4.2)$$

Non Linear Combinations

When we have a classification problem, and our learner outputs discrete class label rather than a real-valued number, a widely used combination rule is non-linear combination among the labels predicted by each member of the ensemble. Kittler and Alkoot theoretically analyzed the relationship the sum and vote fusion strategies. They found that when the errors in out estimate of posterior probability of a class have a normal distribution, sum always performs vote, whereas for heavy tail distributions, vote may outperform sum.

An alternative to the static combination methods discussed so far is to choose a single predictor from the ensemble, dependent on what input pattern is received. The DCS-LA (Dynamic Classifier Selection with Local Accuracy) algorithm by Woods uses estimates of local accuracy in input space to select a single classifier to respond to a new pattern, their experiments establish DCS-LA as robust technique for combining estimators that have individual accuracies. The DCS-LA unlike conventional methods where combinations take place at the output end, selects the classifier to be used at the input end. Fuzzy Integration method for combination output from multiple neural networks is discussed in subsequent sections.

Performance of Combination of neural networks**:**

The performance of a neural network is strongly influenced by two factors. i) number of training samples for each pattern ii) the network size. Classically one starts with an intelligent guess of a neural network size, feeds it with a set of training samples of each pattern to convergence. In principle, it is possible to learn the patterns of any complexity if the training database is quite adequate. If the network size is less than an optimum, it then memorizes the training data poorly, for test set. Several schemes have been suggested for improving the generalization in literature. Of these, schemes based upon collective decision by combining multiple classifiers and possibly multiple sources are well suited for remote sensing data classification. The collective combination tends to improve the interpretation about a given scene that may not fully realizable with one classifier/or with one source.

In the neural network field, several methods for creation of ensembles of neural networks making different errors have been investigated. Such methods basically lie on varying the parameters related to the design and the training of neural networks. In particular, the main methods in the literature can be included in one of the following categories.

In fist method an ensemble of nets can be created by varying the initial random weights, from which each network is trained. In this the nets forming the ensemble exhibit different architectures. Different net types (e.g multi layer perceptrons, radial basis function networks, and probabilistic neural networks) can be used to create the ensemble members.

In second method, an ensemble of nets can be created by training each network with a different learning set. This can be done in a number of different ways. For example, sampling the training data to obtain different learning sets, using learning sets extracted from different data sources (e.g sensors).

Patridge experimentally compared the capabilities of the above methods to create error independent networks. He found the following ordering: varying the net type, varying the training data, varying the net architecture and varying the initial random weights. He thus concluded that varying the net type, varying the training data are two best ways for creating ensembles of networks making different errors. Similar conclusion were shared by other researchers.

The creation of effective neural network still remains an art. Recently researchers have started to investigate the problem of engineering design of neural network ensembles. Proposed approaches can be classified into two main design strategies

1. The direct strategy
2. The over produce and select strategy

The first design strategy is aimed to generate n ensemble of error-independent nets directly. On the other hand, the over produce choose strategy is based on creation of initial large set of nets and the subsequent choice of the subset of most error-independent nets.

5. CONCLUSIONS

It is concluded that by using neural network ensembles there will be considerable improvement in the classification of the remote sensed images. We can also study various methods and have an idea to ensure the error diversity among neural networks in an ensemble. There is also a possibility of combining of different neural networks in an ensemble. The concept of fuzzy measure and fuzzy integration in combining the outputs of the neural networks in an ensemble can also to be considered. The Stabilty problems of Neural networks and convergence problems can also to be minimized by suitably adjusting the learning / training phases of neural networks.

REFERENCES

[1] Blois, M. S., Clinical judgement and computers. *New Eng. J. of Med.* 1980; 303(4): 192-197.

[2] Baxt W.G., Application of artificial neural networks to clinical medicine. *Lancet 1994;346:* 1135-38.

[3] Breiman L., Heuristics of instability in model selection, Technical Report No. 416, Statistics Department, University of California at Berkeley, 1995.

[4] Breiman L., Bagging Predictors, Machine Learning 24 (1996) 123-140.

[5] Breiman L., Combining Predictors, in Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems, A.J.C. Sharkey (ed.) pp 31-50, Springer, London, 1999.

[6] Carney J.G., Cunningham P., The NeuralBAG algorithm: Optimizing generalization performance in bagged neural networks, to be presented at 7f *European Symposium on Artificial Neural Networks,* Bruges (Belgium), 21-23 April 1999.

[7] Carney, J., Cunningham, P., (1999) Confidence and prediction intervals for neural network ensembles, *in proceedings ofIJCNN'99, The International Joint Conference on Neural Networks,* Washington, USA, July, 1999.