Supervised pattern recognition

Ivan E. Villalon-Turrubiates
Instituto Tecnológico y de Estudios Superiores de Occidente (ITESO)
villalon@iteso.mx

Word Count: 3060

Abstract:
Pattern recognition is the scientific discipline that focuses on the classification of data, objects or, in general terms, patterns into categories or classes. To achieve this goal, the methodology uses the extraction of information from the data observation, learn to recognize the different patterns contained within the data and make a decision based on the category of the patterns. This involves supervised classification methods, which are based on external knowledge of the area within the sample to be studied, and therefore, requires some a priori information before the chosen classification algorithm can be applied. The supervised methods are implemented using two main paradigms, statistical algorithms, and neural algorithms. The statistical approach uses parameters that are derived from sampled data in the form of training classes. The neural approach does not rely on statistical information derived from the sample data but is trained directly on the sample data.

Key Words: pattern recognition; pattern theory; classification methods; supervised classification

The term "pattern" refers to an example or model that can be copied. Moreover, it can be interpreted as an imitation of a model as well. A pattern can describe different types of objects that are contained within the physical and abstract worlds by means of its interrelation of data, events, concepts, and context. Therefore, they can be perceived even with certain level of distortion, ambiguous or incomplete information, or with the presence and affectation of noise. As an example, if the letter "P" is distorted, it could be a "P" or an "R", or even more, a totally different letter. However, if the letter appears on the word "OARK", then it has a higher possibility to be the letter "P", but if it appears on the word "OOAD" then it is more likely to be the letter "R", this means that the main context used by the word in the experiment defines the correct letter even with its distortion, which explains the crucial importance of the context in a pattern recognition process (Shih 2010).

The pattern recognition process can be described with two main tasks, one of them is the development of a decision rule that will be based on human knowledge, this is known as the learning process. Afterwards, the rule that has been previously defined for the decision-making process and which will be applied to an unknown pattern is defined as the classification process. The learning process is concerned with the study of recognition mechanisms of patterns by human and other living organisms. The classification process deals with the development of the theory and technique needed for an automatic recognition task.

The application of the pattern recognition theory can be grouped as:

1. Human-machine interaction and communication: This group is focused on automatic speech recognition, optical character recognition systems commonly known as OCR, knowledge extraction from natural language, speech understanding techniques, image and video processing including artificial vision, among others.

2. Biomedical applications: This group focuses on electrocardiogram (ECG) analysis, electroencephalograms (EEG) analysis, electromyography (EMG) analysis, X-Ray analysis, image processing techniques, audio and signal processing methodologies, among others.
3. Physics: This group focuses on high-energy physics, astronomy, quantum mechanics, among others.
4. Criminology studies: This group focuses on the analysis of fingerprints, handwriting, speech, sound, and photography as evidence.
5. Natural resources management: This group focuses on the main study of agriculture, hydrology, forestry, geology, environment, cloud formation, urban assessment, aerosols, and particles, among others.
6. Military applications: All the previous groups are included, adding detection of nuclear hazard, missile guidance and detection, radar and sonar analysis, LIDAR analysis, target identification, and traceability, among others.
7. Industrial applications: This group focuses on computer-aided design and manufacturing, computer simulation in testing and assembly of products, automatic control for quality measurements, ergonomics, among others.
8. Archaeology and anthropology studies: image processing, radar and sonar analysis, resource management, urban dynamics and knowledge extraction from remote sensing imagery, among others.

As a definition, pattern recognition is the scientific discipline that focused on the classification of data, objects or, in general terms, patterns into categories or classes. To achieve this goal, the methodology is focused on the extraction of information from the data observation, learn to recognize the different patterns contained within the data (learning process), and make a decision based on the category of the patterns (classification process). This involves capabilities of unsupervised learning, supervised learning and adopting minimum risk strategies.

A basic process for pattern recognition is statistically based and focuses on the statistical properties of the patterns (i.e. probability densities). The pattern is represented in a feature space, where the features are object's characteristics usually expressed with numeric data. The classes or categories are groups of patterns having similar values on its features and according to a given metric, each class has its own model that is its own mathematical description.

The main goal of the pattern recognition process is to hypothesize the model that is describing each population class, to process the data to eliminate noise (values that are not providing useful information for the classification process), and to choose the model that better corresponds and assign it to the class described by that model.

The unsupervised classification technique can be used to partition a large set of data into groups, called clusters, so that the data points in a group are similar to each other, while those in distinct groups are not similar to those in the other groups. This process is usually selected when the datasets are not originally grouped and the quantity of groups within the data is unknown, however, the datasets can be partitioned into distinct groups based on the similarities or distances among the data points.

The supervised classification technique can be used when the classification is based on a given number of groups and the objective is to assign any new data point to one of these groups. For this process, the number of groups is well-known and some data from each group is also known and used to characterize each particular group, therefore, once the new unclassified data points are analyzed the known information allows to define a classification rule to assign those point to a particular group.

Classification Process

The main process of classification used by pattern recognition methodologies consists of two stages. The first stage is the recognition of object's categories. As an example, considering the context of remote sensing images of the land surface these categories could include woodlands, water bodies, grassland and other land cover types, depending on the geographical scale and nature of the study. The second stage in the classification process is the labelling of the entities to be classified, for the considered example, these labels are numerical so that a pixel that is recognized as belonging to the class 'water' may be given the label '1', 'woodland' may be labelled '2', and so on. Therefore, the process of image classification requires to perform the following steps:

1. To determine a priori the number of categories in which the image will be described. This is the previously described learning process.
2. To assign the numerical labels to the pixel on the basis of their properties using a decision rule. This is the previously described classification process.

Moreover, a set of grayscale values for a single pixel measured in a number of spectral bands contained within the remote sensing image is known as the pattern, and those spectral bands that define the pattern are called features. Therefore, a pattern is the set of measurement on the chosen features for the individuals that are to be classified. The classification process is considered as a pattern recognition methodology for this example, this is, the identification of the pattern associated with each pixel position within the image in terms of the characteristics of the objects or materials that are present at the corresponding point.

Supervised Pattern Recognition

Supervised classification methods are based on external knowledge of the area within the sample to be studied. Unlike the unsupervised methods, the supervised methodologies require some input from the user (or a priori information) before the chosen algorithm can be applied. This input may be derived from different methodologies, as an example considering remote sensing images, the a priori information could be provided by fieldwork, air photo analysis, reports or from the study of known maps for the area of interest. The supervised methods are implemented using two main paradigms, statistical algorithms, and neural algorithms. The statistical approach uses parameters that are derived from sampled data in the form of training classes, as an example, minimum and maximum values of the features, or the mean values of the individual clusters, or the mean and variance matrixes for each class. The neural approach does not rely on statistical information derived from the sample data but is trained directly on the sample data, this is an important characteristic of neural methods because make no assumptions concerning the frequency distribution of the data. The statistical methods are known as parametric because they use statistical parameters that are derived from the training data, whereas neural methods are known as non-parametric (Mather 2004).

The supervised classification method requires prior knowledge of the number of classes and, for the statistical classifiers, some descriptive information from those classes is also required. Considering a remote sensing image, the statistical characteristics of the classes that are to be estimated from the training sample pixels depends on the method that will be used. The simple parallelepiped method requires estimates of the extreme values on each feature for each class, while the k-means or centroid method need estimates of the multivariate means of the classes. The maximum likelihood algorithm requires estimates of the mean vector and variance matrix of each class. On the other hand, the neural classifiers operate directly on the training data but are strongly influenced by the mis-identification of training samples and by the size of the training datasets. The mis-identification of an individual training sample pixel may not have much influence on a statistical classifier, but the impact on a neural method could be considerable.

The a priori knowledge of the number and statistical characteristics of the classes should be reliable and this is a matter of crucial importance for a correct and accurate process. The accuracy of a supervised classification analysis will depend on two factors:

1. The representativeness of the estimates of both the number and the statistical nature of the information classes present in the dataset.
2. The degree of departure from the assumptions upon which the classification technique is based.

These assumptions may vary from one technique to the other, however, the most sophisticated technique will have the most stringent assumptions. Moreover, the validity of statistical estimates depends upon two factors, the size and the representativeness of the sample. The sample size is not simply a matter of "the bigger the better", is related to the number of variables whose statistical properties are to be estimated, the number of those statistical properties, and the degree of variability present in the class.

As an example, considering a remote sensing image, in the case of a single variable and the estimation of a single property such as the mean or the variance, a sample size of 30 pixels is usually held to be sufficient. For the multivariate case, the size should be at least $30p$ pixels per class, where p is the number of features (e.g. spectral bands). The classification accuracy tends to improve as the sample size increases (Dobbertin et al. 1996), therefore, depending on the computer performance available for the user the sample size can be increased in order to obtain a more robust result. On the other hand, neural-based classifiers appear to improve its performance compared with the statistical approach for small training sets, though better results were achieved when training set size was proportional to class size and variability (Blamire 1996).

Training samples are normally located by fieldwork or from air photograph or map interpretation (for the sample of remote sensing images), and their positions on the image found either by visual interpretation, carrying out a geometric correction on the image, or by geo-positioning of the image to be classified. Therefore, a map coordinate will be converted to the corresponding image in column and row coordinates, the location of the image of a training sample whose map coordinates are known is a relatively simple matter, provide that the geometric transform is accurate.

The minimum sample size previously specified is valid only if the individual members of the training sample are independent, as would be the case if balls were drawn randomly from a bag. Generally, adjacent pixels are not independent, as an example for a remote sensing image and considering a pixel identified as "forest", its neighbor pixel would also be a member of the class "forest". If those pixels were statistically independent there would be an equal chance that the second pixel was a member of any other of the candidate classes, irrespective of the class to which the first pixel was allocated. The correlation between nearby points in an image is known as spatial auto-correlation. It follows that the number of pixels in a sample is an over-estimate of the number of fully independent pieces of information in the sample if the pixels making up the training sample are auto-correlated, which may be the case if a block of pixels is selected rather than scattered. The consequence of autocorrelation is that the use of the standard statistical formulas to estimate the means and variances of the features, and the correlations among those features, will provide biased results. Correlations between spectral bands derived from spatially auto-correlated training data will, in fact, be underestimated and the accuracy of the classification will be reduced as a result. The variance matrixes are considerably greater when computed from randomly selected pixels within a class rather than from contiguous blocks of pixels from the same class. The degree of auto-correlation will depend upon the natural association between adjacent pixels, the pixel dimension and the effects of any data pre-processing. The degree of auto-correlation can be calculated by taking sequences of pixels that are spaced several units (predefined) apart and plotting the correlations between a set of pixels and its first, second, third and subsequent nearest neighbors in the form of a correlogram (Basu et al. 1974).

Another source of error encountered in the extraction of training samples is the presence in the sample of atypical values. For instance, one or more vectors of pixel measurements in a given training sample may be contaminated in some way, therefore, the sample means vector and variance matrix for that class will be in error. There are some methodologies to detect those atypical values using estimators of the mean vector and variances matrix which are robust, this is, they are not influenced by those atypical values (Campbell 1980).

The performance of a classifier is usually evaluated using measures of classification accuracy. These accuracy measures use a test set of known data that is collected using the same principles as those described above for the training dataset. The training data is used to calibrate the classifier and test data being used for validation (Muchoney et al. 2002).

Statistical Classifiers

Three algorithms are described in this section, all require that the number of categories (classes) to be specified in advance and that certain statistical characteristics of each class are known:

1. The parallelepiped or box classifier, where a parallelepiped is considered as a simple geometrical shape consisting of a body whose opposite sides are straight and parallel. A parallelogram is a two-dimensional parallelepiped. To define such a body all that is required is an estimate for each class of the values of the lowest and highest pixel values in each band or feature used in the analysis. Pixels are labeled by determining the identifier of the box into which they fall.
2. The centroid classifier, which uses information about the location of each class in the *p*-dimensional Cartesian space defined by the *p* bands (features) to be used as the basis of the classification. The location of each class in the *p*-space is given by the class mean or centroid.
3. The maximum likelihood classifier, which also uses the mean as a measure of the location of the center of each class in the *p*-space and, in addition, makes use of a measure summarizing the disposition or spread of values around the mean along each of the *p* axes of the feature space.

Neural Classifiers

One of the best image interpretation system is the combination of the eyes and the brain. Signals are received by two sensors and are converted into electrical impulses and transmitted to the brain, which interprets them in real-time producing labeled and three-dimensional images of the view field. The brain is composed of a large number of simple processing units called neurons. The number of neurons is of the order of a hundred billion (Greenfield 1997), each neuron is connected to around 10,000 other neurons (Beale et al. 1990). The neurons are connected together in complex ways so that each neuron receives as input the results produced by other neurons, and in turn outputs its signals to other neurons.

The artificial neural networks are based on this model of the brain by building sets of linked processing units in analogy with the neurons of the brain and using these to solve problems. Each neuron is a simple processing unit which receives weighted inputs from other neurons, sums these weighted inputs, performs a simple calculation on this sum such as thresholding, and then sends this output to other neurons. The learning process is accomplished by providing training samples and comparing the actual output of the artificial neural network with the expected output. If there is a difference between the two, then the weights associated with the connections between the neurons are adjusted so as to improve the chances of a correct decision and diminish the chances of the wrong choice being made, and the training step is repeated. This simple model is called single layer perceptron and it can solve classification problems for pattern recognition applications.

See Also:
saseas0025
saseas0099
saseas0108
saseas0147
saseas0165
saseas0395
saseas0403
saseas0553
saseas0562
saseas0630

References

Basu, J.P. and Odell, P.L. 1974. “Effects of intraclass correlation among training samples on the misclassification probabilities of Bayes’ procedure.” *Pattern Recognition* 6: 13–16.
Beale, R. and Jackson, T. 1990. *Neural Computing: An Introduction*. Bristol: Adam Hilger.
Blamire, P. 1996. “The influence of relative sample size in training artificial neural networks.” *International Journal of Remote Sensing* 17: 223–230.
Campbell, N.A. 1980. “Robust procedure in multivariate analysis. I: Robust covariance estimation.” *Applied Statistics* 29: 231–237.
Dobbertin, M. and Biging, G.S. 1996. “A simulation study of the effect of scene autocorrelation, training size and sampling method on classification accuracy.” *Canadian Journal of Remote Sensing* 22: 360–367.
Greenfield, S. 1997. *The Human Brain: A Guided Tour*. London: Weidenfeld and Nicholson.
Mather, P. 2004. *Computer Processing of Remotely-Sensed Images: An Introduction*. Boston: John Wiley & Sons.
Muchoney, D. and Strahler, A.H. 2002. “Pixel and site based calibration and validation methods for evaluating supervised classification of remotely sensed data.” *Remote Sensing of Environment* 81: 290–299.
Shih, F. 2010. *Image Processing and Pattern Recognition: Fundamentals and Techniques*. Hoboken: John Wiley & Sons.