Combining of Image Classification With Probabilistic Neural Network (PNN) Approaches Based on Expectation Maximum (EM)

Wawan Setiawan[1), and Wiweka[2)

[1)Computer Science, Indonesia University of Education

[2)LAPAN Republic of Indonesia

Abstract

This paper presents the design of classifiers with neural network approach based on the method used Expectations Maximum (EM). The decision rule of Bayes classifier using the Minimum Error to the classification of a mixture of multitemporal imagery. In this particular, the multilayer perceptron neural network model with Probabilistic Neural Network (PNN) is used for nonparametric estimation of posterior class probabilities. Temporal image correlation calculated with the prior joint probabilities of each class that is automatically estimated by applying a special formula that is algorithm expectation maximum of multitemporal imagery. Experiments performed on two multitemporal image is the image of the Saguling taken at two different time. Based on experimental results on two test areas can be shown that the average accuracy rate PNN classifier is better than the Back Propagation (BP), and the Expectation Maximum (EM) increase the ability of classifiers. Multinomial PNN classifier by applying the maximum expected to have a consistent recognition capability for multitemporal imagery, and also consistent for each object class category. The proposed classification methodology can solve the problem multitemporal efectively.

Keywords: *Probablistic Neural Networks, Expectation Maximum, Multitemporal Images.*

1. Introduction

Application of neural network to pattern classification have been extensively studied in the past many years. Various kinds of neural network architecture including multilayer perceptron (MLP) neural network, radial basis function (RBF) neural network, and self organizing map (SOM) neural network have been proposed. Because of ease of training and a sound statistical foundation in Bayesian estimation theory, probabilistic neural network (PNN) has become an efective tool for solving many classification problems. However, there is an outstanding issue associated with PNN concerning network structure determination, that is determinating the network size, the location of pattern layer neurons as well as the value of the smoothing parameter. As a matter of fact, the pattern layer of a PNN often consist of all training samples of which many could be redundant. Including redundant samples can potentially lead to large network structure, which in turn induces two problems. First, it would result in higher computaional overhead simply because the amount of computation necessary to classify an unknown pattern is proportional to the size of the network. Second, a consequence of a large network strcture in that the classifier tends to be oversinsitive to the training data and is likely to exhibit poor generalization capacities to the unseen data. On the other hand, the smoothing parameter also plays a cricial role in PNN classifier, and an appropriate smoothing parameter is often data dependent. The two problems mentioned above have been realized by some researchers and some for reduction of training, samples have been proposed. The vector quantization approach was employed to grouph training samples and find cluster centers to be used for PNN. The probability density function of a PNN was approximated by a small number of component densities and the parameters of the components were estimated from the training set by using a Gaussian clustering self organizing algorithm. The clustering technique of the restricted Coulomb energy paradigm was used to find cluster centres and associated weights corresponding to the number of samples represented by each cluster. Basically, all the above mentioned PNN reduction algorithm are based on the clustering approach. Since the classification error has not been used directly in the process of neuron selection, these algorithm can be classified into the category of unsupervised learning.

To estimate a single-date, multivariate, conditional probabilities, need to fuse multitemporal data (for example, to estimae $P(w_i/X_i)$, have to fuse the multitemporal data in I_1). In general, the definition a common statistical model of multitemporal data t_1 and t_2 images may be quite complex. Therefore, adopted a nonparamteric technique. In particular, utilized multilayer perceptron neural networks which, if properly trained by PNN modified algorithm, provide estimates of posterior class probabilities; such estimates can be optimized according to a predefined criterion. In this case, adopted the minimum mean square error (MSE) criterion.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

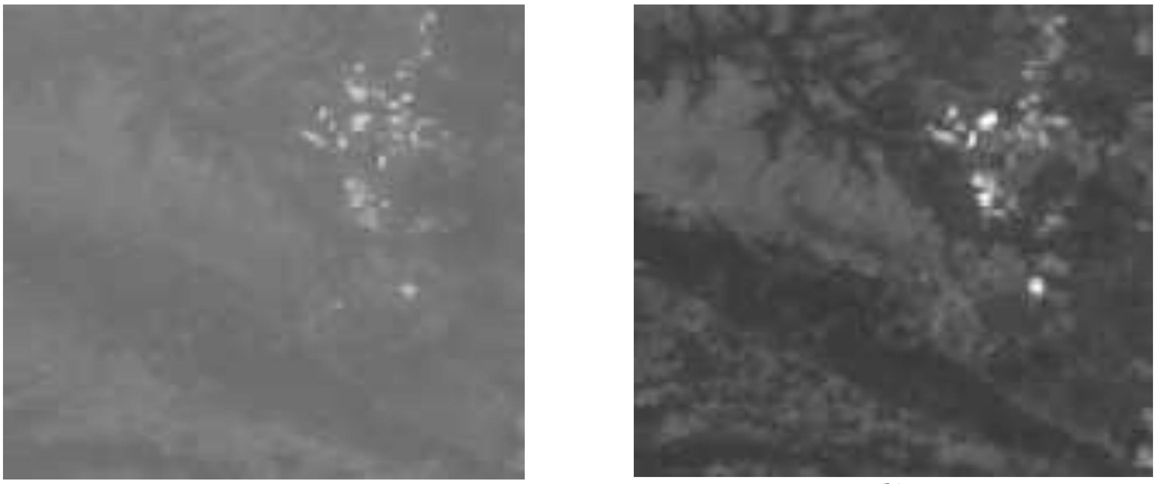
To this end, two PNNs (one for the date t_1, the other for the date t_2) need to be trained separately on two training sets, for which the ground truth must be available. The two training sets may be defined independently of each other.

As a result of training phase, the optimal values of the internal parameters of the neural networks (i.e., the so called weights and biases) are obtained. At this point, each neural network can be used to compute the estimate of the posterior class probability at the corresponding date (output of neural network), given the feature vector X_i (input of PNN).

2. Methodology and Experiment

2.1 Dara of Experiment

A pair multitemporal (Landsat TM) images of Saguling area was used for experiments. The Landsat TM images of Saguling was recorded in July 4^{th}, 1987 (t_1) and July 9^{th}, 1994 (t_2). These images are shown in Figure 1. They also show the location of sample selection used for experiment. The methodology of data processing is illustrated as a block diagram in Figure 2. The available ground truth was used to prepare the training sets (utilized to train neural networks and to estimate prior single-class single-date probabilities) and the test sets (utilized for performance evaluation and comparison).



(a) (b)

Figure 1. (a) Lansat TM image of Saguling area recorded in 1987;
(b) Lansat TM image of Saguling area recorded in 1994.
(Source of data: LAPAN Republic of Indonesia)

Input data consists of images of a region at two different times. Ground-truth is used for training (the need for training the neural network and estimation of prior probability of single-class single-date) and testing (for performance evaluation purposes). Two feature vectors X_1, and X_2 (relative time t_1 and t_2 respectively) consists of 12 texture features derived from co-occurrence matrix (Murni, 1997).

Two neural network in use to estimate the posterior probability P (w_i/X_1) and P (v_j/X_2). Architecture of the fully-connected single hidden layer neural network defined in an independent second respectively for t_1 and t_2 (Lee *et. al.* 1987).

On the application of the EM algorithm selected convergence parameter $\varepsilon = 0.001$. Joint Probability for testing is obtained at the last iteration on the training of the EM algorithm. Temporal correlation between images with an average estimate of EM on joint probability be done to reduce significantly to the classification error on a set of data (Dempster, 1997).

Table 1 provides the classes and the related numbers of pixels includes in the training and test sets.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

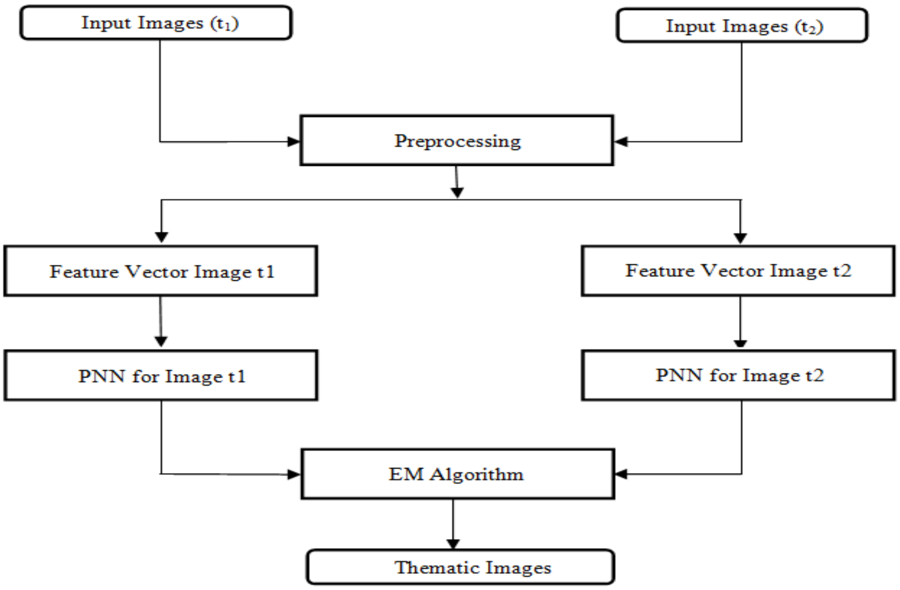


Figure 2. Classification Scheme

Tabel 1. Training and Testing Data.

Number Pixels of Each Class	
Training	Testing
1.800	2.700

Tabel 2. The Atributes of Probabilistic Neural Network (PNN).

Parameter of PNN	**Data**
Network Model	Basic
Kernel	Gauss
Allowable Error	0,001
Sigma Low	0,003
Sigma High	5,0
Sigma Tries	5,0

Tabel 3. The Atribut of Expectation Maximum (EM).

Parameter of EM	**Data**
Error Criteria	0,001
Size of JP	12 x 12
Number of Epoch	288

2.2 Feature Extraction

Feature extraction using a statistical calculation based on a count of second degree involving the relationship of certain characteristics (gray level) between a single-pixel image at a certain position (neighborhood). Included in this level include calculating the probability of occurrence of pixels with gray level g_1 adjacent pixels gray level to g_2, (co-occurrence matrix), the counting of gray level different between two pixels with a certain distance and direction (semivariogram).

One aspect of the texture associated with the spatial distribution and spatial dependence of gray levels of certain areas. One form can be expressed dependence of the probability of appearance together (co-occurrence) of pixels with gray level gray pixels g1 with g_2 gray level. Co-occurrence gray level can be represented in a matrix C in which every element of his c_{ij} is the value of fugnsi P (i, j, d, a). The function P (i, j, d, a) can be read as the probability of occurrence of pixels with gray level gi is at a distance d from the pixels with gray level of g_j in the direction of a direction that could constitute an argument with a possible value of 0^o, 45^o, 90^o, and 135^o. The steps are as follows:

a. Determine the number of gray levels that differ in that image, then sorted from small to large.
b. Form A matrix of size k x k where k is the number of gray levels, where the elements a_{ij} of his stated number of occurrences of pixels with gi kebuan level appears adjacent to the pixels with gray level of g_j in the k.≤ i, j ≤d irection 0^o where 1 1≤ i, j ≤ k.
c. Co-occurrence matrix C is formed by dividing each element of A by n the sum of all elements mantrik A.
d. The next step is to perform statistical calculations on the C matrix. If c_{ij} is an element of the i-th row and j-th column of size k x k matrix C, then some of the characteristic textures that can be obtained through the calculation bellow.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

1) **Anguler Second Moment (ASM)**

$$\sum_{i=1}^{k}\sum_{j=1}^{k} c^2_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

ASM indicates homogeneity or diversity of textures. The more homogeneous texture, the smaller the size of the matrix, but the value of each elemannya the greater, so the value of ASM for more homogeneous texture.

2) **Entrophy**

Entrophy is the degree of randomness (randomness) of the texture. Entropy shows the greatest value if all elements of the same c_{ij}.

$$-\sum_{i=1}^{k}\sum_{j=1}^{k} c^2_{ij} \log c^2_{ij} \quad \ldots\ldots\ldots\ldots\ldots (2)$$

3) **Elemen Difference Moment of m^th^ Order**

Moment of the second level (second order elements Difference Moment) is often referred to as the contrast of textures.

$$\sum_{i=1}^{k}\sum_{j=1}^{k} (i-j)^m \; c_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

4) **Inverse Elemen Difference Moment of m^{th} Order**

The moment of first degree (First Order Moment Difference Inverse Element) is often referred to as the homogenity of texture.

$$\sum_{i=1}^{k}\sum_{j=1}^{k} \frac{c_{ij}}{(i-j)^m} \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

5) **Maximum Probability (max c_{ij})**

The larger the value proabilitas maximum, the dominant appearance of a gray level gi appears adjacent to the gray level in the image of g_j.

6) **Minimum Probability (min c_{ij})**

The larger value of maximum probability, the more dominant appearance of a gray level g_i appears adjacent to the gray level in the image of g_j.

7) **Avarage Probability (avr c_{ij})**

The larger value of average probability, the more dominant appearance of a gray level g_i appears adjacent to the gray level in the image of g_j.

8) **Mean Probability (mean c_{ij})**

The larger value of mean probability, the more dominant appearance of a gray level g_i appears adjacent to the gray level in the image of g_j.

9) **Median Probability (med c_{ij})**

The larger value of median probability, the more dominant appearance of a gray level g_i appears adjacent to the gray level in the image of g_j.

10) **Modus Probability (mods c_{ij})**

The larger value of modus probability, the more dominant appearance of a gray level g_i appears adjacent to the gray level in the image of g_j.

11) **Correlation**

$$\sum_{i=1}^{k}\sum_{j=1}^{k} (i-m)\,(j-m)\;\; c_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

m is the average probability of occurrence (the average value of all the elements of matrix C).

12) **Cluster Shade**

$$\sum_{i=1}^{k}\sum_{j=1}^{k} (i-j-2m)^2 \; c_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(6)$$

m is the average probability of occurrence (the average value of all the elements of matrix C).

2.3 Probabilistic Neural Network (PNN)

Neural network are frequently employed to classify patterns based on learning from examples. Different neural network paradigms employ different learning rules, but all in same way determine pattern statistics from a set of training samples and then classify new patterns on set basis of these statistics.

Current methods such as back propagation use heuristic approaches to discover the underlying class statistics. The heuristic approaches usually involve many modifications to the system parameters that gradually improve system performance. Besides requiring long computaion times for training, the incremental adaptation approach of back propagation can be shown to be susceptible to false minima. To improve upon this approach, a classification method based on estableshed statistical principles was sought.

It will be shown that the resulting network, while similar in structure to back propagation and differing primarily in that the sigmoid activation function is replaced by a statistically derived one, has the unique feature that under certain easily met conditions the decision boundary implemented by the probabilistic neural network (PNN) asymtotically approaches the Bayes optimal decision surface.

To understand the basis of the PNN paradigm, it is useful to begin with a discussion of the Bayes

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

decision startegy and nonparametric estimators of probability density functions. It will then be shown how this statistical technique maps into a feed-forward neural network structure typified by many simple processors ("neurons") that can all function in parallel.

The accurasy of the decision boundaries depends on the accuracy with which the underlying PDFs are estimated. Construct a family of estimates of f(X),

$$fn(X) = \frac{1}{n\lambda} \sum_{i=1}^{n} \varpi\left(\frac{X - X_{Ai}}{\lambda}\right) \quad \text{.....................(7)}$$

which is consistent at all points X at which the PDF is continous. Let XA_1, …, Xa_i, …, X-a_n be dependent random variables identically distributed as a random variable X whose distribution function F(X) = P[x≤ X] is absolutely continues. Parzen's conditions on the weighting function ϖ(y) are :

$$\sup_{-\infty<y<+\infty}|\varpi(y)| < \infty \quad \text{...................... (8)}$$

where sup indicates the supremum.

$$\int_{-\infty}^{+\infty} |\varpi(y)|\,dy < \infty, \quad \text{........................ (9)}$$

$$\lim |\varpi(y)| = 0, \quad \text{....................... (10)}$$

and

$$\int_{-\infty}^{+\infty} \varpi(y)\,dy = 1, \quad \text{...................... (11)}$$

In eqn (1), λ = λ(n) is chosen as a function of n such that :

$$\lim_{y \to \infty} \lambda(n) = 0. \quad \text{.......................... (12)}$$

Proved that the estimate fn(X) is consistent in quadratic mean in the sense that :

$$E|\,fn(X) - f(X)|^2 \to 0 \quad \text{as } n \to \infty \quad \text{......................... (13)}$$

This definition of consistency, which says that the expected error gets smaller as the estimate is based on a larger data set, is particularly important since it means that the true distribution will be approached in a smooth manner.

The Parzen's results can be extended to estimates in the special case that the multivariate kernel is a product of univariate kernels. In the particular case of the Gaussian kernel, the multivariate estimates can be expressed as :

$$fA(X) = \frac{1}{(2\pi)^{p/2}\sigma^p} \frac{1}{m} \sum_{i=1}^{m} \exp[-(X - XAi)^t (X - XAi)/2\sigma^2] \quad \text{................... (14)}$$

where

i = pattern number

m = total number of training patterns

XA_i = i^{th} training pattern from category θA

σ = smoothing parameter

p = dimensionality of measurement space.

Note that fA(X) is simply the sum of small multivariate Gaussian distributions centered at each training sample. However, the sum is not limited to being Gaussian. It can, in fact, approximate any smooth density function.

There is the striking similarity between parallel analog networks that classify patterns using noparametric estimators of a probability density function (PDF) and feed-forward neural networks used with other training algorithms (Swain, 1978).

2.4 Classification Of Mixture Data

Consider the problem of classifying a geographical area acquired at two time t_1 and t_2, respectively. Each data set may contain images derive from different time. Assume that all the images of the two data sets refer to the same ground area, and that they are coregistered and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

appropriately transformed into the same spatial resolution.
In general, the spatial and temporal contextual informatioan plays an important in the classification process. One of the main purposes of this research is to assesss the potentiablities of the technique, it is proposed to estimate prior joint class probabilitis, which are related to the temporal context of the two data sets. Therefor, for simplicity, research focus on the temporal context only, and do not explicitly consider the spatial context. Futhermore, it is assumed that, for each pixel of one data set, all the temporal contextual information is conveyed by the spatially corresponding pixel of the other data set. This seems a reasonable assumption for the current procedure by which consider only two acquisition dates and disregard the spatial context.
Characterize the above pair of temporally corelated pixel (x_1, x_2), x_1 being a pixel of the image data set I_1 and x_2 the spatially corresponding pixel of the image data set I_2, by a pair fetaure vectors (X_1, X_2). Each feature vector X_i is obtained by stacking together the measures provided by the available sensors, as is done for the stacked vector approach. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{M1}\}$ be the of possible land-cover classes at time t_1, and let $N = \{v_1, v_2, \ldots, v_{M2}\}$ the set of possible land-cover classes at time t_2.
It can consider two different types of classification: 1) the "compound classification" of each pair pixel (x_1, x_2), which involves finding the "best" pair of classes (wi, vj) to be assigned to each pair of pixel; or 2) the classification of the pixel of one of the two image data sets by utilizing the information contained in both image data sets I_1 and I_2.
As a classification strategy, need to adopt the Bayes rule for minimum error and apply it to the type of classification, i.e., the "compound classification" of (x_1, x_2), $x_1 \in w_m$ and $x_2 \in v_n$, that

$$P(w_m,v_n/X_1,X_2) = \max_{wi,vj} \{P(w_i,v_j/X_1,X_2)\}. \qquad (15)$$

For the second type of classification, i.e. if the image data set I_2 is to be classified, the Bayes rule become $x_2 \in v_n$ so that :

$$P(v_n/X_1,X_2) = \max_{vj} \{P(v_j/X_1,X_2)\} \qquad (16)$$

The two classification problems, as well as their solutions, are tightly linked. In this research, I shall facus only on the former, i.e., the compound classification of the two image data sets. Under the conventional assumption of class-conditional independence in the time domain, can be written :

$$p(X_1,X_2/w_i,v_j) = p(X_1/w_i)\, p(X_2/v_j) \qquad (17)$$

Consquently, can realice the search for the maximum with :

$$\max_{wi,vj} \left\{ \frac{P(wi/X_1)P(vj/X_2)P(wi,vj)}{P(wi)P(vj)} \right\} \qquad (18)$$

In general, the above assumption may lead to suboptimal solution. For example, when the ground cover assosiated with a given pixel does not change between t_1 and t_2 [i.e., $wi \equiv vj$ in (10)], it is likely that some properties of the ground, wich contribute to determining the values of the sensorial measures, may be saved between the two acquisition dates. This implies the correlation between X_1 and X_2 that is not taken into account. However, adopted the above assumption as it allows a significant simplification of the problem.
The a priori class probabilities P(wi) and P(vj) are estimated from the two training sets by computing the relative frequency of each class. The estimations of the remaining terms (i.e., the posterior class probabilities and the prior joint probabilities) are worth being considered more deeply.

2.5 Estimation Joint Probability With Expectation Maximum (EM) Algorithm

The EM algorithm constitutes a general approach to an iterative compotation of maximum-likelihood estimates of parameters when there is a many-to-one mapping from an underlying distribution to the distribution governing an observation. Such an algorithm is

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012
ISSN (Online): 1694-0814
www.IJCSI.org

particularly popular usefull in the estimating the component of mixture distribution.
The general formulation of the EM algorithm consists of two major steps: an expectation step and a maximization step. The expectation is computed with respect to the unknown underlying variables, using the current estimates of the parameters and conditioned upon the observations. The maximization step provides new estimates of the parameters. These two steps are iterated until convergence.
An important aspect of the EM algorithm concerns its convergence properties. It is possible to prove that, at each iteration, the estimated parameters provide an increase in the likelihood function until a local maximum is reached. Despite the fact that convergence can be ensured, it is impossible to ensure that the algorithm will converge to the global maximum of the likelihood (only in specific cases is it possible to guarantee the convergence to the global maximum).
A detailed description of EM algorithm and of the related theoritical aspect is beyond the scope of this research. Refer to the literature for an in-depth analysis of such an algorithm.

2.6 Estimation Of Prior Joint Probabilities

In this case estimated only to the prior joint probabilities of the classes, assuming no need to update the estimates of the posterior probabilities and of the a priori probabilities of classes during the successive iterations.
The probabilities P(wi,vj) are regarded as the element of the matrix JP, (of size M_1 x M_2, which is computed by maximizing the following:

$$L(JP)=\prod_{q=1}^{S}\left(\sum_{wn\in\Omega}\sum_{vm\in N}P(wn,vm)P(X_1^q,X_2^q/wn,vm)\right) \quad \text{......................} (19)$$

where S is the total number of pixels to be classified and X_k^q is the qth pixel of the image I_k. it is possible to prove that the recursive equation to be used to estimate P(wi,vj) by the maximizing (20) is :

$$P_{k+1}(wi,vj)=\frac{1}{S}\sum_{q=1}^{S}\frac{P_k(wi,vj)P(X_1^q,X_2^q/wi,vj)}{\sum_{wn\in\Omega}\sum_{vm\in N}P_k(wn,vm)P(X_1^q,X_2^q/wn,vm)} \quad \text{……………......…} (20)$$

where $P_k(wi,vj)$ is the iterative joint probability estiamte at the k^{th} iteration. Such estimates are initialized by assigning equal probabilities to each pair of classes.

$$P_0(wi,vj)=\frac{1}{M_1M_2} \quad \forall w_I\in\Omega,\ v_j\in N \quad \text{…..……..…..…… } (21)$$

Under the hypothesis made in approach, it is possible to prove that (21) can be written as:

$$P_{k+1}(wi,vj)=A_{ij}\sum_{q=1}^{S}\frac{P_k(wi,vj)P(wi/X_1^q)P(vj/X_2^q)}{\sum_{wn\in\Omega}\sum_{vm\in N}\frac{P_k(wn,vm)}{P_k(wn)P_k(vm)}P(wn/X_1^q)P(vm/X_2^q)} \quad \text{…...................} (22)$$

where

$$A_{ij}=\frac{1}{S\ \ P(wi)P(vj)} \quad \text{............………....} (23)$$

The algorithm is iterated until convergence. Convergence is reached when the maximum difference between the estimates at two succesive iterations is bellow a threshold. More precisely, the stop criterion is defined by the following :

$$\max_{wi,\ vj}\left\{\left|P_{k+1}(wi,vj)-P_k(wi,vj)\right|\right\}<\varepsilon,\quad wi\in\Omega,\ vj\in N \quad \text{..................} (24)$$

Where $\varepsilon \in [0,1]$, the estimates of P(wi,vj) obtained at at convergence are then applied to the compound classification rule.

2.7 Joint Classification

Two multitemporal remote-sensing images acquired ata times t_1 and t_2 on the same area on the ground are examined. Let us consider couples of pixels made up of a pixel of the multitemporal image acquired at time t_1 and a

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.
IJCSI
www.IJCSI.org

spatially corresponding pixel of the multitemporal image acquired at time t_2: let such pixels be characterized by the d-dimensional feature vector X_1 and X_2, respectively. Let $\Omega = \{ w_1, w_2, \ldots, w_n\}$ be the set of possible land-cover classes at time t_1, and let $N = \{v_1, v_2, \ldots, v_m\}$ be the set possible land-cover classes at time t_2. A land-cover change in the considered couple of pixels is the detected if the two classes w_i and v_j, to which such pixels are assigned, are different.

If we disregard contextual information in t he spatial domain, i.e., if we classify each couple of pixels independently of any other on the basis only of its feature vectors X_1 and X_2, the optimal classification, in the sense of minimum error probability, is given by the Bayes rule for the case of compound classification problems. Such a rule requires that the couple of classes (w_i,v_j), given the observed feature vectors X_1 and X_2:

$$\max_{wi,vj} \left\{ P\left(\frac{wi.vj}{X_1 X_2.} \right) \right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (25)$$

The couple of classes (w_i,v_j) that provides the maximum is the name that provides the following maxima:

$$\max_{wi,vj} \left\{ \frac{p\left(\frac{X_1 X_2}{wi.vj} \right) P\left(\frac{vj}{wi} \right) p(wi)}{p(X_1 X_2)} \right\} \Leftrightarrow \max_{wi,vj} \left\{ p\left(\frac{X_1 X_2}{wi.vj} \right) P\left(\frac{vj}{wi} \right) p(wi) \right\} \ldots\ldots\ldots (26)$$

where the term $p(X_1,X_2)$ can be neglected, as it is independent of w_i and v_j.
Both equation above involve the estimations of n x m functions which are defined in a (2 x d) dimention space. These estimation could be carried out by using a set of training pixels ("training set"). Unfortunately, in real situations, it is difficult to have suitable training set available, as a large number of training pixels for each possible combination of classes w_i and v_j are required. In order to simplify the estimation of such functions, we introduce the following hypothesis. Consider the feature vector Xi (i=1,2), related to time ti, be composed of a signal component Si and of a noise component N_i.

$$X1 = S1 + N1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (27)$$

and

$$X2 = S2 + N2. \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (28)$$

Assume that the signal Si depends only on the land-cover class at time ti, and that the noise Ni depends only on the land-cover class at time ti and possibly on Si (as ocurs, for example, for multiplicative noise in SAR images). Under this hypothesis, the probabilitic dependence of the classes at the two times, and one can write :

$$p\left(\frac{X_1, X_2}{wi,vj} \right) = p\left(\frac{X_1}{wi} \right) p\left(\frac{X_2}{vj} \right) \ldots\ldots\ldots\ldots\ldots (29)$$

By substituting, and applying some transformations, obtained that the following maximum can be used in the decision rule:

$$\max_{wi,vj} \left\{ \frac{P\left(\frac{wi}{X_1} \right) P\left(\frac{vj}{X_2} \right) P\left(\frac{vj}{wi} \right)}{p(vj)} \right\} \quad \ldots\ldots\ldots\ldots\ldots (30)$$

Under the abive-defined hypothesis, to perform the compound classification of two multitemporal remote-sensing images nedd to estimate the a priori probabilities $P(v_j)$ of the classes at time t_1 and t_2, the single-date, multivariate, conditional probabilities $P(w_i.X_1)$ and $P(v_j.X_2)$ at the two times, and the probabilities of transitions $P(v_j/w_i)$.

3. Results and Discussion

The size of the images is 350 x 350 pixels. Secondary features were generated based on the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

co-occurrence model. Number of object classes of Saguling is four including water body, open area, vegetation, and villages. Number of samples each class is 4.500 pixels, 40% of them (1.800 pixles) are used for training and 60% of them (2.700 pixels) are used for testing.

Tabel 4. Classifications Use
Backpropagation Neural Network (BP).

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy- PA (%)	
Water Body	2.562	67	62	-	95,21	
Open Area	71	2.529	34	85	93,01	
Vegetation	67	67	2.533	97	91,64	
Villages	-	33	71	2.518	96,03	
User's Accuracy - UA (%)	94,89	93,67	93,81	93,26	**Overall Accuracy – OA=93.91**	$\overline{PA}$ =**93.97** $\overline{UA}$ =**93.94**

Misclassification of Error water body and villages is relatively much less than the class of open areas and vegetation. Classification of water body and villages conflict with two other classes, while the class of open area and vegetation conflict with all existing class. Classifier is better for the classification of water body and villages. Class of water body interpreted by BP as an open classroom area and the vegetation is almost the same. Similarly villages classes interpreted incorrectly as an open classroom area and vegetation. Two state error is expected because windows co-occurrence size too small so that the same element can be demonstrated by the conflict classes.

Tabel 5. Classifications Use
Probabilistic Neural Network (PNN).

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy- PA (%)	
Water Body	2.609	57	42	-	96,34	
Open Area	51	2.579	-	55	96,05	
Vegetation	40	64	2.573	57	94,11	
Villages	-	-	85	2.588	96,82	
User's Accuracy - UA (%)	96,63	95,52	95,30	95,85	**Overall Accuracy – OA=95.82**	$\overline{PA}$ =**95,83** $\overline{UA}$ =**95.82**

PNN classifier showed better results than BP classifier, all classes of maximum conflict with two other classes. Individually PNN can reduce class conflict. Cases of conflict of water body and villages is still the same but the frequency can be reduced. Misclassification due to the substance of the object is in a different class as an example of the substance of water and green spaces in the vegetation may result in pixels categorized as a class of water body and villages. Then the substance of vegetation and open spaces in villages are not interpreted as villages.

Tabel 6. Classifications Use
Back Propagation Neural Network (BP) with Expectarion Maximum (EM).

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy- PA (%)	
Water Body	2.596	51	38	-	96,15	
Open Area	55	2.587	-	58	96,15	
Vegetation	49	62	2.583	51	96,15	
Villages	-	-	79	2.591	96,15	
User's Accuracy - UA (%)	96,15	95,81	95,67	95,96	**Overall Accuracy – OA=95.90**	$\overline{PA}$ =**95.91** $\overline{UA}$ =**95.90**

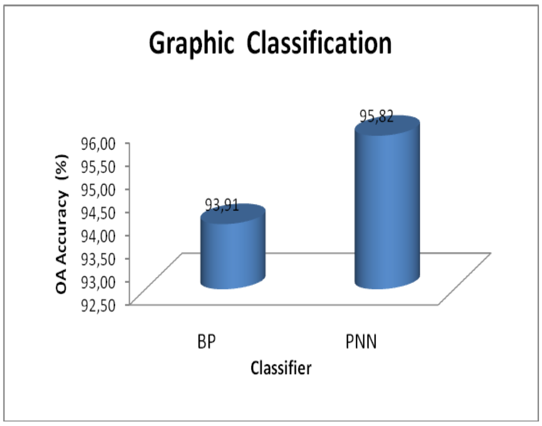


Figure 3. BP and PNN Accuracy.

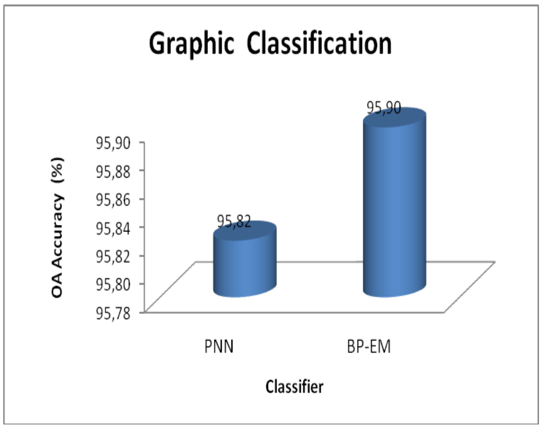


Figure 4. PNN and BP-EM Accuracy.

The Expectation Maximum (EM) can enhance the ability of BP classifier, reduce the number of class conflict, all classes only conflict with a maximum of two other classes. The combination of BP-EM also is better than individual PNN approximately 0.1%. Expectation maximum may serve to give certainty that the substance of the decision classifying objects that are not logically be included in the class of the closest object.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Tabel 7. Classifications Use
Probabilistic Neural Network (PNN) with ExpectatrionMaximum (EM).

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy - PA (%)	
Water Body	2.648	24	25	-	98,18	
Open Area	8	2.625	-	42	98,13	
Vegetation	44	51	2.637	20	95,82	
Villages	-	-	34	2.638	98,73	
User's Accuracy - UA (%)	98,07	97,22	97,81	97,70	Overall Accuracy – OA=97.67	$\overline{PA}$ =97.72 $\overline{UA}$ =97.70

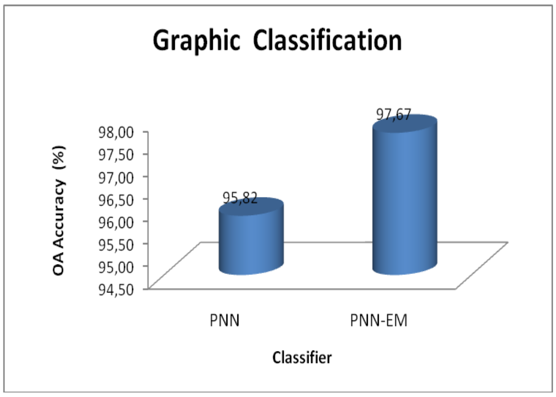


Figure 5. All Classifier Accuracy

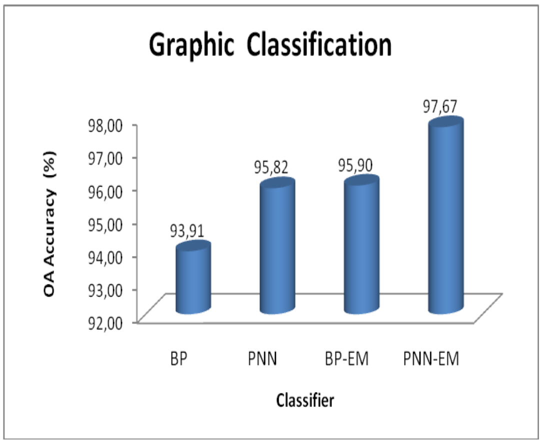


Figure 6. PNN and PNN-EM Accuracy

PNN compound with EM can improve object recognition multitemporal imagery approximately 2% more from all other schemes. Misclassification of water body as vegetation still remains difficult even with EM. Misclassification of water body as open area is reduced by EM as highest capability. Open area and villages still be object most difficult to classify with EM even misclassification can be reduced.

4. Conclusion

Concerning the multitemporal aspect, we assumed that, for simplicity, the temporal correlation between the two data sets can be taken into account by the prior joint probabilities of classes at two dates. Multitemporal data was then performed by means of probabilistic neural networks, which provide nonparametric estimates of posterior class probabilities on the basis of single-date.

In addition to the flexibility provided by nonparametric techniques, the use of neural networks offers the general advantages of every neural network approach: instrinsic parallelism, adaptability to data, and robustness to noise and errors o training data . in particular for the neural model we adopted, no general rules exist to define the neural network topology and establish the procedur of the training process; moreover, it is difficult to interpret the network behavior.

To increase accuracy and training time, we uses PNN. One of principle advantages of the PNN pradigm is that it is very much faster than the well-known back propagation paradigm for problems in which the incremental adaptation time of back propagation is a significant fraction of the total computation time. Classification accuracy was roughly comparable, back propagation produced 93,91% where as PNN produced 95,82% over all accuracy.

The main innovation is EM algorithm for estimation of the prior joint probabilities of classes. Prior joint probabilities are usually chosen manually by a human expert on the basis of a prior kwonledge derived from the characteristic of the geographical area considered and from the time interval between acquisitions. The advantage of the EM algorithm consists in the possibility of computing the estimates of joint probabilities directly from the data set to be classified and in a fully automatic way. This overcomes the drawbacks resulting from the need for a human intervention and from the dependence on the accuracy of a rpior knowledge.

Acknowledgements

The research was founded by the Higher Education (DIKTI) through a national strategic research program in 2011.This research done with the help and good cooperation of all parties. For that, thanks to Chairman of the Computational Intelligence and Image Processing Laboratory of Computer Science UPI. Specialy thanks to Prof. Aniati Murni from University of Indonesia. Hopefully the results can be a contribution in the efforts to optimize the development of land used and cover of Indonesia.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

References

Aniati Murni, " Metodologi interpretasi citra inderaja multitemporal dan multisensor berdasarkan klasifikasi uniform", Disertasi Pascasarjana UI, Jakarta, 1997.

Bin Tian, et al., "Temporal Updating Scheme for Probabilistic Neural Network with Application to Satellite Cloud Clasiffication", IEEE Trans. On Neural Network, Vol. 11 No. 4, July 2000.

Byeungwoo J. et. al, "Decision Fusion Approach Multitemporal Classifications ", 1999.

Fausett, L., "Fundamentals of Neural Networks", Prentice Hall, 1994.

Goldberg, D. E., "Genetic Algorithm in Search, Optimization, and Machine Learning", Addison Wesley, 1989.

H. M. Kalayeh and D. A. Landgrebe, "Utilyzing multitemporal data by stochastic model", IEEE Trans. Geosci. Remote Sensing, vo. GE-24, pp. 792-795, Sept. 1986.

H. Solberg, A. K. Jain, and T. Taxt, " Multisource classification of remotely sensed data: Fuion of landsat TM and SAR images", IEEE Trans. Geosci. Remote Sensing, vo. 32, pp. 100-113, Jan. 1997.

J. A. Benediktsson, and P. H. Swain, "A method of statistical multisource classification with a mechanism to weihgt the influence of the data sources", In Proc. IEEE Trans. Geosci. Remote Sensing, pp. 517-520, July 1989.

J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approach versus statistical method in classification of multisouce remote sensing data", IEEE Trans. Geosci. Remote Sensing, vo. 28, pp. 540-552, July 1990.

Jose R. G. A., et al, "Applications of the EM Algorithm to the Analysis of Life Length Data", Appl. Statist. Vo. 44, No. 3 1995.

Josef K. et al, "On Combining Classifiers ",IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, March. 1998.

L. Bruzzone and S. B. Serpico, " An iterative technique for detection of land-cover transitions in multitemporal remote-sensing images", IEEE Trans. Geosci. Remote Sensing, vol. 35, pp. 858-867, July 1997.

L. Bruzzone, C. Conese, F. Maselli, and F. Roli, " "Multisource classification of complex rural areas by statistical and neural network approaches", Photogramm. Eng. Remote Sens., vol. 63, no. 5, pp. 523-533, May 1997.

L. Bruzzone, D. F. Prieto, and S. B. Serpico, " A Neural –Statistical Approach to Multitemporal and Multisource Remote Sensing Image Classification", IEEE Trans. Geosci. Remote Sensing, vol. 37, pp. 1350-1358, May 1999.

Migual A. et al, "Mode-Finding for Mixtures of Gaussian Distributions", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, Nov. 2000.

P. Dempster, N. M. Laird, and D. B. Rubin, " Maximum likelihood from incomplete data via the EM algorithm", J. R. Stat. Soc., vo. 39, no. 1, pp. 1-38,. 1997.

P. H. Swain, "Bayesian classification in a time-varyng environment", IEEE Trans. Syst. Man. Cyber., vol. SMC-8, pp. 880-883, Dec. 1978.

T. Lee, J. A. Richard and P. H. Swain, "Probabilistic approach for multisource data analysis", IEEE Trans. Geosci. Remote Sensing, vol. GE. 25, pp. 283-293, May 1987.

Wawan Setiawan is Doctor Biographies in the Department of Computer Science at the University of Indonesia. He works at Computer Science, Indonesia University of Education. His Researches interests include Computational Intelligence, and ICT For Education. His works mainly focus on ICT applications. For years later he developed systems computer based to support educational process.

Wiweka is Doctor Biographies in the Department of Computer Science at the University of Indonesia. He works at LAPAN Republic of Indonesia. His Researches interests and mainly focus on include Image Processing. For years later he developed systems computer based to support land used recognition.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.