Land Cover Classification Based On MODIS Imagery Data Using Artificial Neural Networks

Arthur Stepchenko
Riga Technical University.
Address: 2 Daugavgrivas Str., Riga, LV-1048, Latvia,
Ventspils University College.
Address: 101a Inzenieru Str., Ventspils, LV-3601, Latvia

Abstract. Remote sensing has been widely used to obtain land cover information using automated classification. Land cover is a measure of what is overlaying the surface of the earth. Accurate mapping of land cover on a regional scale is useful in such fields as precision agriculture or forest management and is one of the most important applications in remote sensing. In this study, multispectral MODIS Terra NDVI images and an artificial neural network (ANN) were used in land cover classification. Artificial neural network is a computing tool that is designed to simulate the way the human brain analyzes and process information. Artificial neural networks are one of the commonly applied machine learning algorithm, and they have become popular in the analysis of remotely sensed data, particularly in classification or feature extraction from image data more accurately than conventional method. This paper focuses on an automated classification system based on a pattern recognition neural network. Variational mode decomposition method is used as an image data pre-processing tool in this classification system. The result of this study will be land cover map.

Keywords: Artificial Neural Networks, Normalized Difference Vegetation Index, Pattern Recognition, Variational Mode Decomposition.

I. INTRODUCTION

Agricultural and forest resource mapping, in which land cover classification is an important element, is mostly performed with remote sensing data. Land cover is the physical and biological cover at the surface of the earth [1]. Examples of land cover classes include water, snow, grassland, forest, scrubland, wetland, asphalt and bare soil. Traditional methods of land cover mapping have been limited to field surveys that are time-consuming and uneconomical with data collected over long time intervals [2]. Remote sensing has been a worthwhile source of information over last three decades in mapping land cover because remotely sensed data have a large geographic coverage, short revisit periods over the same point on Earth and good image quality. It is more economical and faster technique compared to traditional methods of land cover mapping. Remote sensing can be used to obtain land cover information using manual interpretation or automated classification techniques. Usually, the pixels of the remote sensing image are grouped into meaningful and homogeneous land cover classes using digital image classification. Digital image classification is the process where each pixel is assigned the value of land cover class. The objective is to classify each pixel into only one class [2].

Temporal-related features are important for improving land cover classification accuracy using remote sensing data [3]. Vegetation indices calculated from satellite images can be used to obtain information associated with land cover. Vegetation indices is some combination of spectral bands designed to take out a particular property of vegetation [4]. The normalized difference vegetation index (NDVI) is developed for estimating vegetation cover from the red and near infrared band of satellite data. The NDVI is an indicator, which quantifies the amount of green vegetation. The NDVI index is calculated by:

$$NDVI = \frac{NIR - R}{NIR + R}, \qquad (1)$$

where *NIR* represents the spectral reflectance value (i.e. digital number) in near infrared band and *R* represents the spectral reflectance value in red band. Greener and dense vegetation has low red light reflectance values and high near infrared reflectance values, and therefore high NDVI values. The NDVI values are scaled between minus one and plus one, where higher positive values are corresponding to greener vegetation, but low positive values and negative values are corresponding to non-vegetated surface features such as water, barren land, rock, ice, snow, clouds or artificial materials [5].

This paper explores capabilities of using feature extraction method - variational mode decomposition (VMD) to decompose NDVI images. VMD is an enhanced version of the empirical mode decomposition (EMD) algorithm for analyzing non-linear and non-stationary signals, including two-dimensional signals. It adaptively decomposes the signal, e.g. image, into a set of band-limited oscillations called modes of separate spectral bands [6]. These modes or sub-signals can be used as features to make classification using some artificial intelligence or statistical classification technique. In this paper was used artificial intelligence method - pattern recognition neural network. The artificial neural network is distribution-free method that have many advantages in classification tasks including possibility to process complex, noisy, non-linear and non-stationary data [7]. Artificial neural networks are very good at pattern recognition problems and with enough neurons can classify any data with arbitrary accuracy.

II. MATERIALS AND METHODS

A. Study Area

The study site is a Ventspils Municipality that is located in the western part of Courland, Latvia. Its total area is 2472 km2 (Fig. 1). The region plays an important role in agricultural and forest management in Latvia.

Fig. 1. Ventspils Municipality in orange color.

The climate in the study site is determined by temperate climate zone with significant maritime features. Approximately half of the area of Ventspils Municipality is covered by forests, and the other half is covered by grasslands.

B. NDVI Data Set

Multi-temporal, smoothed, geometric and radiometric corrected MODIS Terra NDVI image was used with spatial resolution 250 m and temporal resolution 7 days (Fig. 2). Image was taken in July 2016.

Fig. 2. MODIS Terra NDVI image from this study.

Ground truth data were obtained from The Latvian Geospatial Information Agency. These data include land cover samples in Ventspils Municiplaity.

C. Variational Mode Decomposition

Variational mode decomposition method (VMD) non-recursively decomposes a multi-component signal $f(t)$ into various band-limited monocomponent signals or intrinsic mode functions $u_k(t)$ using calculus of variation [8]. Each mode is regarded as an amplitude-modulated and frequency-modulated (AM-FM) signal and is assumed to have compact frequency support around a central frequency ω_k [9]. VMD tries to find out these central frequencies and intrinsic mode functions centered on those frequencies concurrently using an optimization methodology called alternate direction method of multipliers (ADMM). The original formulation of the optimization problem is continuous in time domain. The constrained variational problem is the following:

$$\min_{u_k,\omega_k} = \{\sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] * e^{-j\omega_k t} \right\|_2^2 \} \quad (2)$$

$$s.t. \sum_k u_k(t) = f(t)$$

where t is a time, j is an imaginary number, e is an exponent, π is a constant (3.14159) and ∂_t is a derivative at time t.

To evaluate the bandwidth of the modes, first the associated analytic signal by means of the Hilbert transform u_k^H is calculated to obtain a unilateral frequency spectrum for each mode. The frequency spectrum of this signal is one sided (exist only for positive frequency) and assumed to be centered on ω_k. The frequency spectrum of each mode (k-th mode) is shifted to the baseband (origin) by multiplying with an exponential and signal now is centered at origin:

$$u_k^M(t) = (u_k(t) + j u_k^H(t)) e^{-j\omega_k t} . \quad (3)$$

The integral of the square of the time derivative (or Gaussian smoothness) of this (3) signal is a measure of bandwidth of the intrinsic mode function.

All the modes can be obtained in the frequency domain. First, Fourier transform is used to obtain original signal in frequency (spectral) domain. In frequency domain at *n+1* iteration *k-th* mode is obtained:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k}\hat{u}_i(\omega) + \hat{\lambda}^n(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} . \quad (4)$$

The term in (4):

$$\hat{f}(\omega) - \sum_{i \neq k}\hat{u}_i(\omega) \quad (5)$$

is a result of Wiener filtering, which makes VMD method robust to noise, and *α* is the balancing parameter of the data fidelity (bandwidth) constraint. At *n+1* iteration *k-th* mode central frequency is updated by:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 \, d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 \, d\omega} . \quad (6)$$

Dual ascent for all *ω≥0* at *n+1* iteration is obtained:

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)) . \quad (7)$$

Iterations continues until convergence:

$$\sum_k \left\| \hat{u}_k^{n+1}(\omega) - \hat{u}_k^n(\omega) \right\|_2^2 / \left\| \hat{u}_k^k(\omega) \right\|_2^2 < \varepsilon , \quad (8)$$

where *ε* is some small number – convergence criterion. When all modes are obtained in frequency domain, inverse Fourier transform can be used to obtain these modes in time domain.

D. Artificial Neural Networks

Artificial neural networks (ANNs) are a form of artificial intelligence, is an information-processing paradigm that inspired by biological nervous systems, such as the brain [10]. ANNs are one of accurate and widely used classification models. Structure of artificial neural networks makes them valuable for a classification task with good accuracy. Artificial neural networks are self-adaptive methods that learn from data [11]. Neural networks learn from examples and can find functional relationships between the data even if relationships are unknown or the physical meaning is not clear [10]. Therefore, ANNs are well suited for classification problems, whose solutions are difficult to obtain, but for which there are enough data or observations.

Artificial neural networks can generalize. After learning from the input data (a sample or pattern), ANNs can often correctly process the early unseen sample even if the sample data are noisy. Neural networks are less sensitive to noise better than most other methods.

Pattern recognition neural network is a feedforward neural network that is based on multilayer perceptron. It consists of three or more neuron layers: one input layer, one output layer and one or more hidden layers (Fig. 3). In most cases, a network with only one hidden layer is used to limit calculation time, especially when the results obtained are satisfactory. All the neurons of each layer (except the neurons of the output layer) are connected by an axon to each neuron of the next layer [12].

An individual neuron receives weighted inputs from previous layers, which are summed in each neuron using a combination function. The result of this combined summation is passed through a transfer function to produce the output of the neuron.
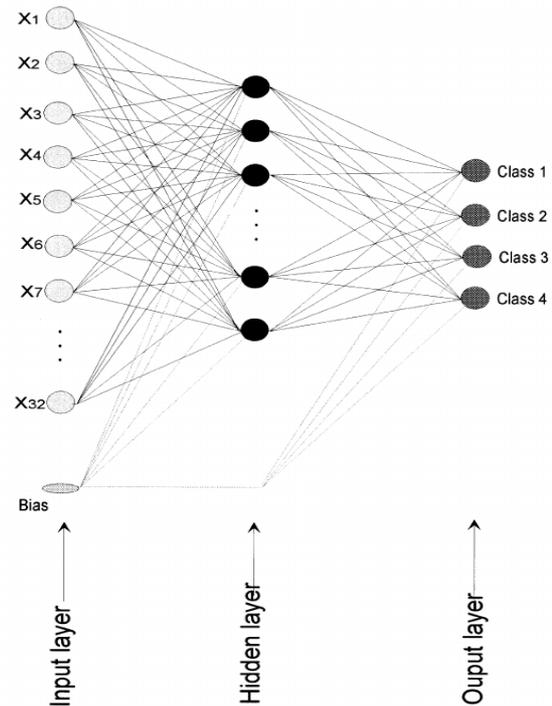

Fig. 3. Multilayer perceptron with one hidden layer.

The combination function and transfer function together constitute the activation function. The most widely used activation function for the output layer in classification task is the softmax function. The softmax function, or normalized exponential function is a generalization of the logistic function that normalize a *K*-dimensional vector *z* of arbitrary real values to a *K*-dimensional vector of real values *σ(z)* in the range (0, 1) where sum of all values (probabilities) is 1. The softmax activation function is given by:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, j = 1,...,K , \quad (9)$$

where *K* is the number of classes.

The sigmoid (logistic) and exponential (hyperbolic) tangent functions are often used as the hidden layer transfer function. To improve the accuracy of the neural network, each data point (each feature) in the input neurons needs to be normalized – rescaled within the range of [−1, 1] or [0, 1]. The multilayer perceptron is trained with error-correction

learning, which means that the desired response (class value) for the system must be known.

To train a neuron with several input variables, $x_1, x_2, \ldots$, corresponding weights $w_1, w_2, \ldots$, and a bias, b the cross-entropy cost (performance) function in case of multi-class classification task for this neuron is defined as [12]:

$$C = -\frac{1}{n}\sum_x \left[y \ln \sigma(z)_j \right], \tag{10}$$

where n is the number of samples in training data, x is the sum over all training samples, and y is the corresponding desired output (observed class). The function returns a result that strongly penalizes outputs that are extremely inaccurate, with very little penalty for completely correct classifications. Minimizing cross-entropy leads to good classifiers.

III. EXPERIMENTAL PROCEDURE

The aim of this experiment is to investigate the capability and accuracy of pattern recognition neural networks in the land cover classification task. In the first stage features were extracted from NDVI image using bi-dimensional variational mode decomposition method The second stage employed pattern recognition neural network for multi-class classification.

A. Feature Extraction

Bi-dimensional variational mode decomposition method was applied to the original NDVI image in order to obtain K sub-images. First, in order to obtain K value, non-parametric bi-dimensional empirical mode decomposition method was applied, that automatically give optimal number of modes. It give K value four. Then with known number of modes, variational mode decomposition was applied with following parameters: number of modes K was four, dual ascent time step τ was 0.1, and convergence criterion was 1e-7. Optimal bandwidth constraint α was found using error-and-trial approach, and it was 4.1. Grater or smaller α lead to modes, the sum of which do not reconstruct original image. Obtained sub-images are given in Table I.

Table I
Extracted sub-images

a) first sub-image	b) second sub-image
c) third sub-image	d) fourth sub-image
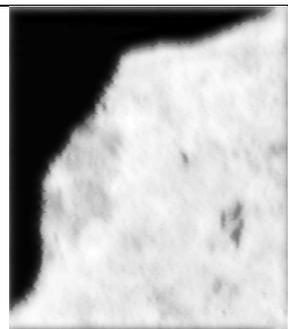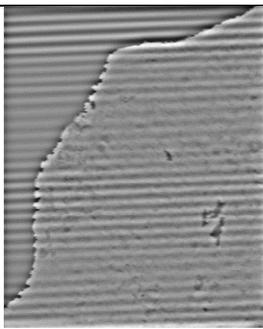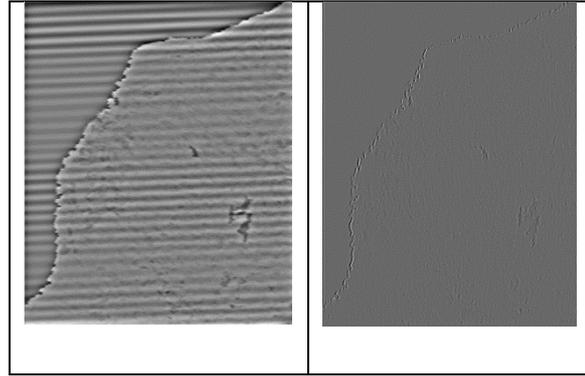
B. Multi-class classification

For each land cover sample in vector map that is located in Ventspils Municipality, corresponding median values calculated from pixel values of original NDVI image and four sub-images were found and input data set for pattern recognition neural network was created. This data set was divided into two sets, training and testing data set by rate 70/30. 70% of the samples (488 samples) were used as a training data set, 30% of the samples (210 samples) were used as a testing data set. Five classes were used: grassland (class 1), forest (class 2), scrubland (class 3), wetland (class 4) and water (class 5).

Pattern recognition neural network (PRNN) model used in this study was trained by scaled conjugate gradient backpropagation training function. Neural network's weights and biases were initialized with small random numbers in [-0.25, 0.25]. The number of network's hidden layers was one. The hyperbolic tangent function and the softmax function are used as activation functions for the hidden and output layers, respectively. The number of epochs that are used to train was set to 10000. As the number of hidden neurons is an important factor that determining the classification accuracy, is required to find an optimal value, but there is currently no theory to determine how many nodes in the hidden layer are optimal. The optimal complexity of PRNN model, that is, the number of hidden nodes, was determined by a trial-and-error approach. In the present study, the number of hidden nodes was progressively increased from 1 to 5. In order to improve neural network generalization ability early stopping technique was used. When the network begins to overfit the data, the global error on the validation set typically begins to rise. When the validation error increased for a 100 epochs in a row, the training was stopped, and the weights and biases at the minimum of the validation error were used. This neural network's configuration was determined experimentally as giving the best results. A program code was written in MATLAB environment. Overall accuracy and confusion matrix were used as measures of correctness in order to determine proposed PRNN model accuracy.

model," *Mathematical Problems in Engineering,* vol. 2014, article ID 201402, 2014.

[11] G. Zhang, B. E. Patuwo and M. Y. Hu, "Forecasting with artificial neural networks: the state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35-62, March 1998.

[12] D. Reby, S. Lek, I. Dimopoulos, J. Joachim, J. Lauga and S. Aulagnier, "Artificial neural networks as a classification method in the behavioural sciences," *Behavioural Processes*, vol. 40, pp. 35–43, 1997.