ISPRS Journal of Photogrammetry and Remote Sensing 146 (2018) 161–181

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Progressively Expanded Neural Network (PEN Net) for hyperspectral image classification: A new neural network paradigm for remote sensing image analysis

Paheding Sidike[a,b,*], Vijayan K. Asari[a], Vasit Sagan[b]

[a] Dept. of Electrical and Computer Engineering, University of Dayton, Dayton, OH 45469, USA
[b] Department of Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO 63108, USA

ARTICLE INFO

Keywords:
Neural network
Hyperspectral image (HSI)
Classification
Machine learning
Remote sensing

ABSTRACT

Hyperspectral image (HSI) has been used for a wide range of applications including forestry, urban planning, and precision agriculture. In recent years, machine learning based algorithms, such as support vector machines, decision trees, ensemble learning, and their variations have shown promising results in HSI analysis. Such methodologies, nevertheless, can lead to insufficient information abstraction in interpreting hyperspectral pixels. In this paper, we propose a novel neural network based classification algorithm, named Progressively Expanded Neural Network (PEN Net), that can effectively interpret hyperspectral pixels in nonlinear feature spaces and then determine their categories. Furthermore, a spectral-spatial HSI classification framework is also introduced to test the generality and robustness of the PEN Net. Experimental results on four standard hyperspectral datasets illustrate that: (1) PEN Net classifier yields better accuracy and competitive processing speed in HSI classification tasks compared to the state-of-the-art methods; (2) Multi-hidden layer based PEN Net generally provides better performance than single hidden layer one; (3) Combination of spectral and spatial features in the PEN Net classifier can significantly improve the classification accuracy by 6–15% compared to the spectral only based HSI classification. This study implies that the proposed neural network architecture opens a new window for future research and the potential for remote sensing image analysis.

1. Introduction

Rapid advances in sensor technology, field robotics, unmanned aerial systems, and computing power have facilitated exponential growth in Remote Sensing (RS) applications. Meanwhile, processing complex, multiscale, and high-dimensional data such as hyperspectral data, has become increasingly difficult for RS community. There exist both numerous opportunities and challenges with broader usage of automated image analytic tools that translate the abundant spectral and spatial data to useful information for decision-making.

Hyperspectral remote sensors collect image data in hundreds of narrow and adjacent spectral bands. The different spectral signatures associated with specific materials being imaged assist in automatic target detection and classification since objects vary uniquely from the natural background in absorbing and reflecting radiation at different wavelengths. In most cases, the targets can be differentiated and identified based on their spectral signature, which provides many practical applications in life science, surveillance, agriculture, forestry, and natural resources management. However, considerable spectral variability and subpixel targets make hyperspectral image (HSI) processing challenging (Manolakis and Shaw, 2002).

The objective of HSI classification is to assign a hyperspectral pixel into an object category that it belongs to, which also termed as thematic mapping. To perform the classification task, spectral libraries or training data and ground truth information are generally required, unless to perform unsupervised image classification. Two major challenges in HSI classification can be summarized into two aspects: (1) The sparseness of the target class implies that insufficient availability of the training data; (2) considerable intraclass variability and interclass similarity introduce difficulty in pattern differentiation. To address these challenges, many pixel-wise machine learning methods have been developed, such as maximum likelihood method (Richards and Richards, 1999), Bayesian estimation models (Landgrebe, 2005), Support Vector Machine (SVM) (Cortes and Vapnik, 1995; Gualtieri and Cromp, 1999), decision trees (Goel et al., 2003), Neural Networks (NN) (Subramanian et al., 1997; Yang, 1999; Hernández-Espinosa et al., 2004), genetic

* Corresponding author at: Department of Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO 63108, USA.
E-mail address: sidike.paheding@slu.edu (P. Sidike).

https://doi.org/10.1016/j.isprsjprs.2018.09.007
Received 30 January 2018; Received in revised form 24 July 2018; Accepted 6 September 2018

algorithms (Vaiphasa, 2003), and kernel-based techniques (Camps-Valls and Bruzzone, 2005). Although these methods have been widely examined and extended for HSI analysis, there still exist challenges and scopes to effectively and efficiently process HSI due to ever-increasing data complexity and volume.

Recently, there has been a growing interest on a biological inspired NN for RS data analysis (Pal et al., 2013; Samat et al., 2014; Yue et al., 2015; Chen et al., 2016; Lv et al., 2016). Particularly, Deep Learning (DL) models, which are capable of learning the discriminative features from input data in hierarchical fashion instead of handcrafting features, have been introduced into the geoscience and RS community with the rapid surge of interest in this topic. Three fundamental DL architectures are Deep Belief Network (DBN) (Hinton et al., 2006), Stacked Auto-Encoder (SAE) (Bengio et al., 2007) and Convolutional Neural Network (CNN) (LeCun et al., 1998). Some representative applications of these models for RS data analysis include image preprocessing (e.g., pan-sharpening) (Huang et al., 2015), target recognition (Chen et al., 2014b), RS image classification (Romero et al., 2016; Hu et al., 2015a) as well as segmentation (Albalooshi et al., 2018). Among these applications, HSI classification using DL has become one of the most active research fields in recent years. Chen et al. (2014c) applied SAE to extract hierarchical features from hyperspectral data in the spectral and spatial domain, while Tao et al. (2015) employed stacked sparse autoencoder for spectral-spatial feature representation and showed its discriminative power for HSI classification. Similarly, the effectiveness of DBN has also been explored for HSI data analysis and the results indicated superior performance of DBN compared to SVM. Although DBN and SAE are capable of extracting robust features and generally achieve better accuracy than traditional machine learning classifiers, they cannot efficiently extract contextual spatial information owing to their inherent network structures (Lee and Kwon, 2017). In contrast, CNN can effectively extract spatial and spectral features with less amount of parameters, thus it has been a more preferable deep model in the current trend of HSI analysis. Hu et al. (2015b) formulated a one-dimensional (1D) CNN structure for pixel-wise HSI classification, and a better performance was achieved compared to SVM and a two-layer neural network. Li et al. (2017) introduced an improved 1D-CNN model for HSI by applying CNN on pixel-pairs to augment the number of training samples and achieved the state-of-the-art accuracy. Although these CNN models yielded higher accuracy than other types of popular classifiers such as SVM, a considerable amount of training time and sufficient data requirement limit their performance.

Another popular biologically inspired NN, named Extreme Learning Machine (ELM) (Huang et al., 2006), has attracted more and more attention of the community in diverse research fields (Maimaitijiang et al., 2017; Sidike et al., 2017; Peng et al., 2013) due to its higher regularization performance at a much faster training speed over the state-of-the-art NNs. Pal et al. (2013) proposed a new kernel based ELM algorithm which provided better classification accuracy than the radial basis kernel function based SVM and ELM methods with notable lower computational cost. Moreno et al. (2014) applied ELM and Optimally Pruned ELM (OP-ELM) for soybean variety classification in hyperspectral images, and showed that the OP-ELM yielded the best and more stable results than the previously reported accuracy when using a single spectral band. Samat et al. (2014) introduced an ensemble ELM, which combines the Bagging-based and AdaBoost-based ELMs, for HSI classification, and yielded better accuracy than ELM and SVM. However, the drawbacks of ELMs are that (1) it typically requires a large number of neurons to reach good accuracy which affects its computational merit, and (2) their applications are limited on the relatively easy databases compared to DL methods, thus the combination of ELM with DL or other innovations on its basic principles will be needed (Tissera and McDonnell, 2016). In this work, we introduce a nonlinear function expansion scheme into multi-hidden layer model in our proposed NN which alleviates the abovementioned limitations of ELMs.

There are two types of information can be utilized for HSI classification: spectral signatures and spatial content. Recent advances in spatial resolution enhancement of HSI attracts researchers' interests in exploiting spatial information. Unlike the classic HSI classification methods which only consider the spectral signature of every pixel, the spatial-feature based approaches represent each pixel by extracting contextual information of that pixel in every spectral band. A study by Camps-Valls et al. (2006), a family of composite kernels which integrates spatial and spectral information is introduced. This method formulates a set of kernel-based classifiers by considering spectral, spatial, and local cross-information in HSI. Pesaresi and Benediktsson (2001) investigated Morphological Profile (MP), which utilizes morphological operation to generate spatial structural features, for HSI classification. Due to its successful performance, the improved versions of MP, such as Extended MP (EMP) (Benediktsson et al., 2005) and Extended multi-Attribute MP (EAMP) (Dalla Mura et al., 2010), were developed. For noise-robust HSI classification, Chen et al. (2014a) employed a multi-hypothesis prediction model to incorporate spatial features and reconstruct HSI. Li et al. (2012) proposed to include spatial information in HSI classification using a multilevel logistic Markov-Gibbs random field prior. Kang et al. (2014) effectively utilized edge-preserving filtering as a probability optimization process to improve the classification output, whereas Chen et al. (2015a) extracted edge features by computing spatial and rotational auto-correlations of local image gradients.

Texture information is another useful spatial feature that can aid in HSI classification. Markov Random Fields (MRFs) can be used to extract texture features since they measure spatial relationship between the central pixels and its neighboring pixels, which have been successful applied in HSI classification (Eches et al., 2013; Tarabalka et al., 2010). Gabor features as a widely used texture descriptor have been explored for HSI classification (Huo and Tang, 2011; Bau et al., 2010). For instance, Huo and Tang (2011) computed a two-dimensional Gabor features in a principal component analysis projected subspace to obtain texture features, while a three-dimensional Gabor filter bank was successfully applied to HSI to capture texture features from specific orientation and scale (Bau et al., 2010). Recently, Local Binary Pattern (LBP) (Ojala et al., 2002) has shown surprisingly good performance in HSI classification (Li et al., 2015a). Specifically, the LBP coded image is first generated for each band in the input HSI, then the LBP histogram for each pixel of interest is computed with its corresponding neighborhood region. Recently, a new deviation of LBP, named Volumetric Directional Pattern (VDP) Essa et al. (2017) extracts texture features from the directional magnitude component of three consecutive bands in HSI, was proposed and demonstrated its promising classification accuracy among the other competitors. Although LBP and VDP significantly contribute HSI classification accuracy, neglect the texture features from the local sign and local magnitude components at various scales. In contrast, a multiscale Completed LBP (CLBP) (Guo et al., 2010) is capable of extracting multi-level textural and structural features from images and it can outperform LBP and VDP, as well as the state-of-the-art spatial features in HSI classification task (Sidike et al., 2016, 2018).

In terms of DL frameworks, the exploitation of spatial information has also been found to be substantially effective for HSI classification. Yue et al. (2015) introduced a 2D-CNN architecture where spatial features were incorporated during the classification framework. Lee and Kwon (2017) explored local contextual interactions in HSI through a multi-scale convolutional filter bank which was used as the initial layer of subsequent CNN pipeline. Considering high-dimensionality of HSI data, Makantasis et al. (2015) presented a randomized principal component analysis approach to reduce spectral dimension of HSI while a CNN model was used to encode spectral and spatial information. Similarly, a balanced local discriminant embedding algorithm was proposed to extract low-dimensionality spectral features and then a 2D-CNN model was used to generate high-level spatial features. One of the major drawbacks of 2D-CNN approaches is that they may not fully