International Journal of Applied Earth Observation and Geoinformation 34 (2015) 103–112

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

A support vector machine to identify irrigated crop types using time-series Landsat NDVI data

Baojuan Zheng[a,*], Soe W. Myint[a,c], Prasad S. Thenkabail[b], Rimjhim M. Aggarwal[c]

[a] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287, United States
[b] United States Geological Survey (USGS), 2255 N Gemini Dr., Flagstaff, AZ 86001, United States
[c] School of Sustainability, Arizona State University, Tempe, AZ 85287, United States

ARTICLE INFO

Article history:
Received 19 December 2013
Accepted 18 July 2014

Keywords:
Crop classification
Landsat
NDVI
Support vector machines
SVM

ABSTRACT

Site-specific information of crop types is required for many agro-environmental assessments. The study investigated the potential of support vector machines (SVMs) in discriminating various crop types in a complex cropping system in the Phoenix Active Management Area. We applied SVMs to Landsat time-series Normalized Difference Vegetation Index (NDVI) data using training datasets selected by two different approaches: stratified random approach and intelligent selection approach using local knowledge. The SVM models effectively classified nine major crop types with overall accuracies of >86% for both training datasets. Our results showed that the intelligent selection approach was able to reduce the training set size and achieved higher overall classification accuracy than the stratified random approach. The intelligent selection approach is particularly useful when the availability of reference data is limited and unbalanced among different classes. The study demonstrated the potential of utilizing multi-temporal Landsat imagery to systematically monitor crop types and cropping patterns over time in arid and semi-arid regions.

Introduction

Arizona has a warm, dry climate with two distinct wet seasons in late summer and winter (Sheppard et al., 2002). Studies show that climate change could results in more extreme weather events, such as prolonged and intense droughts in semi-arid environment (Karl et al., 2009; Overpeck and Udall, 2010). With rapidly growing population and urbanization, and the additional challenges from climate change, Arizona is facing major water-management challenges to meet human needs while maintaining ecosystem health (Gleick and Palaniappan, 2010).

Agricultural sector is a vital part of Arizona's economy; it contributed around $4 billion in value added ("Impacts of Agricultural Production", 2010). However, irrigated agriculture consumes about 68% of total water use in the state – the largest use of freshwater in Arizona (Cousteau, 2012). In the events of droughts and water scarcity, agricultural sector is projected to be impacted the most in terms of water availability in Arizona (CAST, 2009). As such, improving water use efficiency in agricultural sector is critical. Monitoring agricultural water use by crop type at different spatial scales could assist in formulating water resource management plans and policies. Furthermore, accurate site-specific information of crop types is required for detailed assessments of agricultural water use and for other agro-environmental assessments.

The potential to use remote sensing imagery for crop classification over large areas has been broadly investigated and crop phenology is the basis for cropland mapping (Chang et al., 2007; Long et al., 2013; Shao et al., 2010; Turker and Arikan, 2005; Vintrou et al., 2012; Wardlow et al., 2007; Zhong et al., 2014). The Moderate Resolution Imaging Spectroradiometer (MODIS) data has high potential for mapping crops worldwide because of its high temporal and moderate spatial resolution attributes. Using MODIS data to map crop types has been implemented across different parts of the world at a regional level (Vintrou et al., 2012; Wardlow and Egbert, 2008). There are also studies that combine MODIS data and moderate spatial resolution data, such as Landsat and the Indian Remote Sensing Advanced Wide Field Sensor (AWIFS), to discriminate crop types (Thenkabail and Wu, 2012; USDA-NASS, 2013). Other studies used higher spatial resolution imagery, such as Landsat and ASTER data, to differentiate crop types for a less extensive area (Peña-Barragán et al., 2011; Serra and Pons, 2008; Turker and Arikan, 2005). Image selection for crop type mapping largely depends on the extensiveness of the study area, image availability, the cost, and the level of diversity in crop types and management. Although both

* Corresponding author. Tel.: +1 646 750 7087.
E-mail address: bzheng11@asu.edu (B. Zheng).

http://dx.doi.org/10.1016/j.jag.2014.07.002
0303-2434/

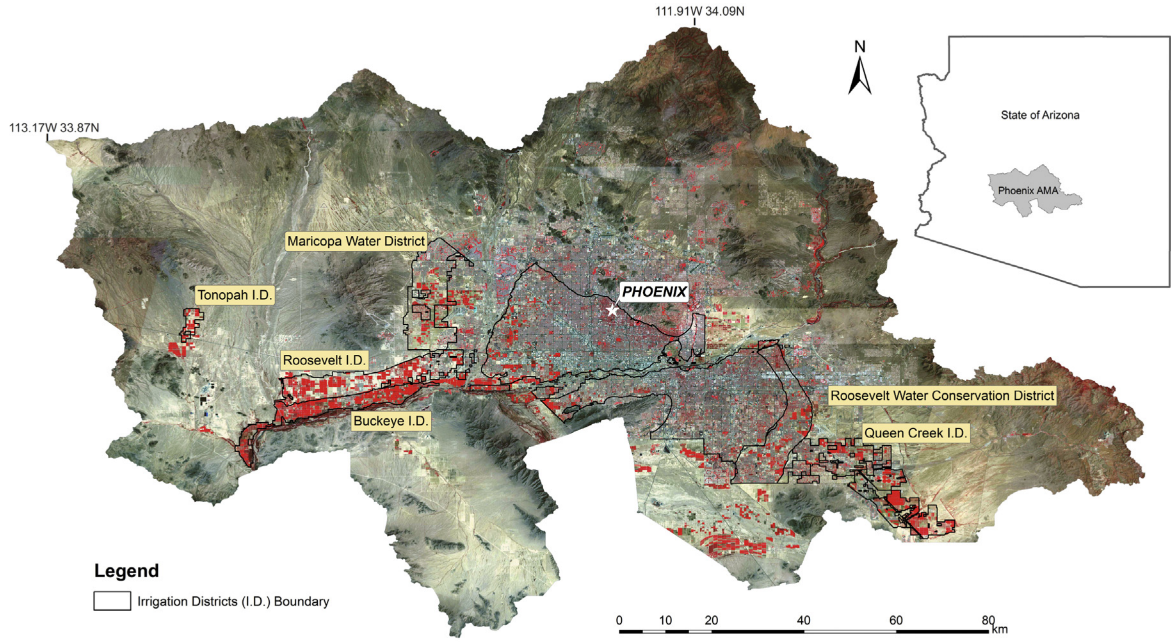

Fig. 1. False color (bands 4, 3, and 2 in red, green, and blue respectively) Landsat imagery of the Phoenix Active Management Area (PHX AMA) overlay with irrigation district boundary. (Note that some agricultural lands do not belong to any irrigation districts.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

MODIS and Landsat data are freely available, MODIS data generally encounter a well-known mix-pixel problem due to its coarse spatial resolution (250–500 m), and the eight-day revisit rate of Landsat 5 TM and 7 EMT + combined cannot provide enough number of cloud-free imagery for most regions in the world. However, the identification of detailed crop types in a complex agriculture management area using Landsat 30 m resolution data can be expected to be more desirable and accurate in comparison to MODIS, especially for many tropical, subtropical regions characterized by smaller agricultural fields. While most regions suffer from limited availability of could-free Landsat imagery, there are plenty in arid, semi-arid, and any dry climate regions. Leveraging this positive aspect, we attempted to effectively map crop types in Arizona, by employing frequent cloud-free Landsat observations.

Given the above background, the study aims to examine the potential of using Landsat time-series Normalized Difference Vegetation Index (NDVI) data to differentiate different crop types in the Phoenix Active Management Area (PHX AMA). The specific objectives of the study are to: (1) characterize seasonal spectral changes (i.e., NDVI temporal profiles) of different crops to assist accurate crop classification in the region; (2) evaluate the effectiveness of support vector machines (SVMs) in discriminating various crop types using a small training dataset; (3) examine if SVMs using Landsat imagery alone are capable of identifying double cropping patterns.

Investigation of these questions will provide the basis for its operational application to survey croplands over time in the PHX AMA region. One of the significant impediments in crop type mapping is the lack of quality training data (Shao et al., 2010). When ideal ground reference data is not available or limited, an alternative approach to acquire reference data is to use prior knowledge of cropping practices and spectral data collected in another year. This study, thus, also provides a direct visual comparison between crop types and NDVI time-series spectral profiles, and information on crop management, to assist acquisitions of high-quality training data. Such information is important for long-term, local, and global cropland mapping efforts (Thenkabail et al., 2012), especially when there is limited availability of ground reference data.

Support vector machine classifier

Support vector machine (SVM) classifier is a non-parametric supervised classification derived from statistical learning theory. SVM was developed in the late 1970s, but its popularity in remote sensing only began to increase about a decade ago (Mountrakis et al., 2011). The underlying theory and detailed mathematical explanation of SVM have been demonstrated in many previous studies (Ben-Hur and Weston, 2010; Foody and Mathur, 2004; Vapnik, 1999). SVM training algorithm maps the training data into higher dimensional space and finds the optimal hyperplanes that separate classes with minimum classification errors. The optimal hyperplanes are positioned using training samples that lie at the edges of class distribution in a feature space. The training samples that define the hyperplane of maximum margin are called support vectors. All the other training samples do not make any contribution to estimate hyperplane locations, and therefore can be discarded (Belousov et al., 2002; Brown et al., 2000). Consequently, it is possible to use SVM to achieve high classification accuracy using a small number of training samples.

Similar to other non-parametric classifiers, SVM does not assume that data is normally distributed for a particular image. This allows SVM to perform better than techniques based on maximum likelihood classification (Bruzzone and Persello, 2009; Dalponte et al., 2008). Previous studies showed that SVM has the ability to generalize to unseen data with a small training dataset (Foody and Mathur, 2004; Plaza et al., 2009; Shao and Lunetta, 2012). Shao and Lunetta (2012) compared SVM to two other non-parametric classifiers, i.e., neural networks (NN), and classification and regression trees (CART). They found that SVM achieved substantially higher classification accuracy than NN and CART using a small training sample size (i.e., 20 pixels per class) (Shao and Lunetta, 2012). SVM is an appealing approach to handle high-dimensional data with a limited training set (Plaza et al., 2009). It was applied to time-series MODIS (Carrão et al., 2008; Shao and Lunetta, 2012) and RapidEye (Löw et al., 2013) data. However, to our best knowledge, SVM has not yet been applied to time-series Landsat data to classify different crop types.