Computers and Electronics in Agriculture 142 (2017) 50–58

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

# Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning

Jason Kane Gilbertson *, Adriaan van Niekerk

*Department of Geography and Environmental Studies, University of Stellenbosch, Stellenbosch, South Africa*

ARTICLE INFO

*Article history:*
Received 23 June 2017
Received in revised form 21 August 2017
Accepted 22 August 2017
Available online 1 September 2017

*Keywords:*
Feature selection
Classification and regression trees
Random forest
Principal components analysis
Pan-sharpening
Machine learning
Crop type mapping

ABSTRACT

This study evaluates the use of automated and manual feature selection – prior to machine learning – for the differentiation of crops in a Mediterranean climate (Western Cape, South Africa). Five Landsat-8 images covering the different crop class phenological stages were acquired and used to generate a range of spectral and textural features within an object-based image analysis (OBIA) paradigm. The features were used as input to decision trees (DTs), k-nearest neighbour (k-NN), support vector machine (SVM), and random forest (RF) supervised classifiers. Testing was done by performing classifications (using all spatial variables) and then incrementally reducing the feature counts (based on importance allocated to features by filters), feature extraction, and manual (semantic) feature selection. Classification and regression trees (CART) and RF were used as methods to filter feature selection. Feature-extraction methods employed include principal components analysis (PCA) and Tasselled cap transformation (TCT). The classification results were analysed by comparing the overall accuracies and kappa coefficients of each scenario, while McNemar's test was used to assess the statistical significance of differences in accuracies among classifiers. Feature selection was found to improve the overall accuracies of the DT, k-NN, and RF classifications, but reduced the accuracy of SVM. The results showed that SVM with feature extraction (PCA) on individual image dates produced the most accurate classification (96.2%). Semantic groupings of features for classification also revealed that using the image bands and indices is not sufficient for crop classification, and that additional features are needed. The accuracy differences of the classifiers were, however, not statistically significant, which suggests that, although dimensionality reduction can improve crop differentiation when multi-temporal Landsat-8 imagery is used, it had a marginal effect on the results. For operational crop-type classification in the study area (and similar regions), we conclude that the SVM algorithm can be applied to the full set of features generated.



## 1. Introduction

Crop maps assist in maintaining the health of an economy's agricultural sector as they are used to update agricultural statistics and aid in yield forecasting (Castillejo-Gonzalez and López-Granados, 2009). An additional benefit of up-to-date crop maps is increased environmental sustainability as these maps are required for modelling greenhouse gas variability in agro-ecosystems (Monfreda et al., 2008). Crop-mapping has traditionally been done using routine field visits. This methodology is costly and can be inaccurate when biased sampling schemes are utilised (Castillejo-Gonzalez and López-Granados, 2009). Remote sensing offers an unbiased, cost-effective, and reliable way of mapping crops at a local, regional, and national scale. Crop discrimination using remotely sensed data is, however, not without challenges. Certain crop types have similar spectral signatures, which makes it difficult to differentiate them from one another when using multispectral imagery. Since different crop types have divergent temporal spectral profiles, the consideration of temporal (phenological) variations between crops have been shown to improve classification accuracies (Castillejo-Gonzalez and López-Granados, 2009). However, some crop types may display intra-class temporal variability (different phenological growth patterns) from farm to farm due to either natural development variation or diverse crop-management decisions made by farmers (Peña-Barragán et al., 2011).

Nevertheless, the value of multi-temporal data for crop discrimination has been demonstrated by Wardlow et al. (2007), Ozelkan et al. (2015), Zheng et al. (2015). Multi-temporal data allows for the generation of a large number of features (variables) for each image acquisition date, which has been shown to substantially

* Corresponding author.
*E-mail addresses:* 16255135@sun.ac.za, jasonkanegilbertson@gmail.com (J.K. Gilbertson).

http://dx.doi.org/10.1016/j.compag.2017.08.024
0168-1699/

improve results (Heinl et al., 2009). However, the use of multi-temporal data often leads to very high feature counts (Lu and Weng, 2007; Heinl et al., 2009). Too many features will lead to the so-called "curse of dimensionality", whereby the performance of classifiers is hampered by the imbalance between training samples and features (Rodriguez-Galiano et al., 2012). This is driven by the problem of sparsity, where training data becomes too sparse to cope with the large volume of feature space brought on by large numbers of variables (Myburgh and Van Niekerk, 2013). Classifiers consequently require an increasing number of training samples as feature dimensionality increases.

Large sets of training samples are not always feasible due to the high costs associated with field visits (Castillejo-Gonzalez and López-Granados, 2009). Another approach to mitigate high dimensionality is to carry out feature extraction and/or feature selection (Guyon and Elisseeff, 2003). Feature extraction involves the replacement of the original data by a new collection of features representing most of the variance of the original data (Benediktsson and Sveinsson, 1997). The most common feature-extraction method is PCA, which transforms the data into a new set of principle components (PCs) that describes the underlying structure of the original dataset (Zhang and Mishra, 2012). Other feature-extraction methods include Tasselled cap transformation (TCT) and the generation of spectral indices. The TCT is a process whereby spectral data from an optical sensor (predominantly Landsat) is compressed into a few bands associated with a scene's physical characteristics while suffering minimal information loss (Huang et al., 2002). Spectral indices are essentially arithmetic operators performed on multispectral imagery (or any additional data), which results in a new composite image (Campbell and Wynne, 2011). Examples include NDVI (normalised difference vegetation index), SAVI (soil-adjusted vegetation index), and EVI (enhanced vegetation index). PCA, TCT, and spectral indices are commonly used when classifying crops with remotely sensed data (Simms et al., 2014; Campbell et al., 2015; Zheng et al., 2015).

Feature selection involves picking a subset of important features from the original dataset to reduce data dimensionality (Guyon and Elisseeff, 2003). The main feature-selection approaches are wrappers, embedded methods, semantic groupings, and filters. A wrapper evaluates various subsets of features during the classification process by making use of the learning algorithm itself (Kojadinovic and Wottka, 2000). The advantages of wrappers include interaction between model selection and feature-subset search, and the capability to take feature dependencies into account. However, wrappers have a high risk of overfitting and are also computationally intensive, as every feature subset proposed by the subset selection measure is evaluated in the context of the learning algorithm (Saeys et al., 2007). Examples of wrappers include recursive feature elimination (Shahi et al., 2016), sequential feature selection (Lagrange et al., 2016), and genetic algorithms (Persello and Bruzzone, 2016).

Embedded feature-selection methods are similar to wrappers as they are also used to optimize the performance of a learning algorithm (Guyon and Elisseeff, 2003). Embedded techniques learn which features contribute the most to the accuracy of the classification while the model is being created. The difference between an embedded approach and a wrapper is that the former method utilizes an intrinsic model-building metric during learning. Examples of embedded methods include L1 (LASSO) regularization and DTs (Huang et al., 2002). Semantic feature selection simply involves the selection of features according to their type or those deemed most important by an expert. Examples include using only spectral features, only indices, only texture features, etc.

A filter is a pre-processing step that is independent of the learning algorithm (Fourie, 2011). This step results in a faster learning pipeline for the feature-selection algorithm (when multiple classifiers are used), but filters tend to not perform as well downstream due to an absence of interaction with the classifier (Kojadinovic and Wottka, 2000). Three popular filter methods used in remote sensing are Jeffries-Matusita distance, CART, and RF (Miner et al., 2009; Rodriguez-Galiano et al., 2012; Hao et al., 2016). CART is a decision-tree machine learning algorithm used for data mining, predictive modelling, and data pre-processing. It uses binary recursive partitioning to grow DTs, while the Gini and Twoing methods search for important relationships and patterns, allowing better insight into data (Breiman et al., 1984). It can be used to create a short list of predictor variables for use with another predictive method (Miner et al., 2009). Yu et al. (2006) used CART for detailed vegetation classification with high spatial resolution imagery and found that it improved classification accuracy. Yu et al. (2006) started with two out of 52 variables and found an increase in overall accuracy with the addition of features from 1 to 27, after which accuracies began to decline. Conrad et al. (2011) analysed the effect of CART feature selection on crop classification accuracy using multi-temporal MODIS imagery. They found that CART was able to improve classification accuracy by up to 7% and ascribed this to the prioritization of segments representing active phases of the different crops' phenological development.

RF is a collection of DTs that form an ensemble learning method for classification or feature selection (Pal and Mather, 2003). Rodriguez-Galiano et al. (2012) assessed the effect of RF feature selection on Mediterranean land-cover classification (including multiple crop classes) with multi-seasonal imagery and texture. They found that feature selection using RF had a positive effect on image classification (overall accuracy increases of up to 10%) and commented that feature selection reduced the effect of the "curse of dimensionality". Hao et al. (2016) utilized RF feature selection for crop classification with multi-temporal MODIS imagery and claimed that the technique allowed the identification of the optimal portion of features required for an accurate discrimination between crop types.

Compared to the traditional pixel-based image analyses (PBIA), OBIA approaches have been shown to produce higher classification accuracies in some cases (Castillejo-Gonzalez and López-Granados, 2009; Yan et al., 2015), while Duro et al. (2015) found that both paradigms produced similar results. In general, OBIA is preferred only if the objects of interest are significantly larger than the pixels of the imagery (Blaschke, 2010).

This study evaluates the use of multi-temporal, object-based supervised classification for the differentiation of crops in a Mediterranean climate (Cape Winelands, South Africa). Five Landsat-8 images were used to generate a large (205) set of features. A small set (159) of fields representing the seven major crops in the region was selected to train and assess the classifiers. The size of the in situ dataset was purposefully limited to evaluate the classifiers' ability to perform with minimal training data (i.e. under sparse training conditions). Filter feature selection (using CART and RF), feature extraction (using PCA and TCT), and thematic feature groupings were applied to the full feature-set to assess whether these techniques improve classification accuracies. The different feature-sets were used to train four machine learning classifiers, namely DT, k-NN, RF and SVM. The classification results are interpreted in the context of finding an operational solution for the production of accurate crop-type maps in the Cape Winelands region.

## 2. Materials and methods

### 2.1. Study area and period

The experiments were carried out in a 1040 $km^2$ area within the Cape Winelands region, South Africa (Fig. 1). The area, which