Some Analytical Techniques for Crop Identification Using Satellite Data

James McBroom
Griffith University & Queensland University of Technology

August 2016

Overview

- Introduction
- Supervised and Unsupervised Classification
- Some Supervised Techniques
- Example Data
- Example Analysis
- Example Results
- Questions

Introduction

- Goal: identify crop type using information from satellites.
- Satellite data: Reflectances, space, time.
- Indices (EVI, NDVI, etc)
- Covariates (meteorology, soil, growing season, etc)
- Ground Truth (Field Observations)

Introduction - Reflectances

- Different materials reflect and absorb solar radiation differently at different wavelengths.
- These spectral "signatures" can be used to identify objects such as bare soil, leafy vegetation etc.
- E.g. Landsat7:

Variable	Description
band1	Blue reflectance
band2	Green reflectance
band3	Red reflectance
band4	Near Infrared (NIR) reflectance
band5	Short–Wave Infrared 1 (SWIR–1) reflectance
band6	Short–Wave Infrared 2 (SWIR–2) reflectance

Introduction - Vegetation Indices

- It has been found that certain combinations of bands (Indices) can be useful in identifying and differentiating types of vegetation.

Acronym	Full Name
CTVI	Corrected Transformed Vegetation Index
EVI	Enhanced Vegetation Index
GEMI	Global Environmental Monitoring Index
LSWI	Land Surface Water Index
MNDWI	Modified Normalised Difference Water Index
MSAVI	Modified Soil Adjusted Vegetation Index
MSAVI2	Modified Soil Adjusted Vegetation Index 2
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NRVI	Normalised Ratio Vegetation Index
RVI	Ratio Vegetation Index
SAVI	Soil Adjusted Vegetation Index
SR	Simple Ratio Vegetation Index
TVI	Transformed Vegetation Index
TTVI	Thiam's Transformed Vegetation Index

Introduction - Covariates

- The addition of extra information can aid the differentiation and identification of vegetation types.
- Soil type, meteorological information, season, etc.
- Spatial information (neighbouring pixels more likely to be similar)
- Temporal information (growth patterns through time)

Introduction - Ground Truth

- Field observations
- Train and verify predictive approaches based on satellite information
- "Truth"? Hopefully. (Assumed)
- Predictive models will only be as good as the information they are trained on.

Classification - Overview

- Two kinds of classification: Supervised and Unsupervised
- Supervised:
  - Group membership is known and used
  - This known membership is used to train a grouping rule
  - Resultant rule is used to classify new, unknown cases.
- Unsupervised:
  - Group membership either unknown or unused
  - Number of groups unknown
  - Can be viewed as exploratory
- Remotely–sensed crop data with ground truth: Supervised Classification.

Classification - Unsupervised

- Examples:
  - Clustering via Mixture Models
  - Hierarchical Agglomerative Clustering
  - K–means Clustering
  - "Ordination" – PCA, MDS
- Will not focus on these approaches.

Classification - Supervised

- Data with known group membership (Ground Truth)
- Random split (e.g. 70/30): Training/Testing
- Misclassification rates – error, or confusion, matrix
- Cross–validation

Supervised Classification Techniques

- Some commonly used supervised techniques
- No one technique is best in all situations
- Strengths and weaknesses

Artificial Neural Networks

- Nonlinear, nonparametric
- “Black Box”.

Support Vector Machines, SVMs

- Work well on binary problems with linear boundary.
- Nonlinear boundaries, multi–class less clear.

Logistic (Binary/Multinomial) Regression

- Parametric, model–based. Hypothesis testing.
- Can be extended to include space and time correlation.
- Potential high computational load.

k–Nearest Neighbour

- Simple and intuitive – membership is majority vote of k nearest neighbours.
- Can be sensitive to local data structures.
- Lazy learner – classification rule made at time of query rather than before.

Decision Trees

- Simple, large amounts of data.
- Can overfit.
- Random Forests [ensemble technique to overcome overfitting issue].

Ensemble Classifiers

- Serial and Parallel versions.
- Can produce more accurate results.
- Currently popular in remote sensing literature.

Space and Time Considerations

- Many (most) supervised classification techniques do not consider spatially or temporally (or both) correlated data.
- Sate–Space models: non-normality, nonlinearity, large amounts of data. Current area of research.

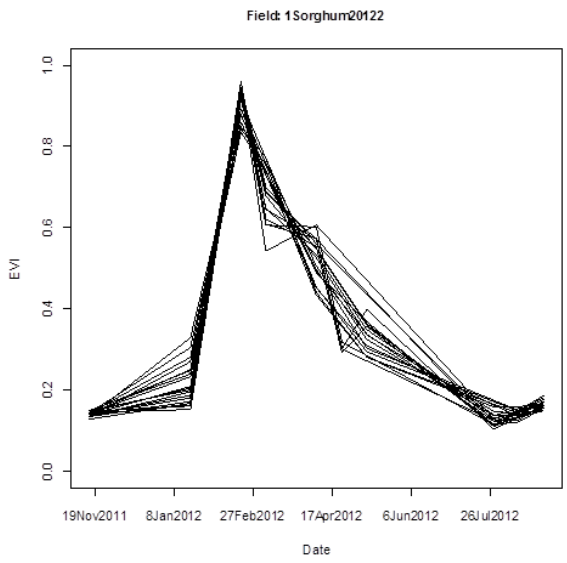
Example Data

Description:

- Data from Landsat 7 satellite: 25m $\times$ 25m pixels, measured approximately every 16 days (Geoscience Australia).
- Spatio–temporally subsetted to correspond to ground–truth data:
  - 1186 locations collected between 2011 and 2012 (QLD DSITIA). Not repeatedly measured.
  - Includes crop type/ground cover but not crop yield.

Example Data Continued

Ground truth grids. Image courtesy of ABS.



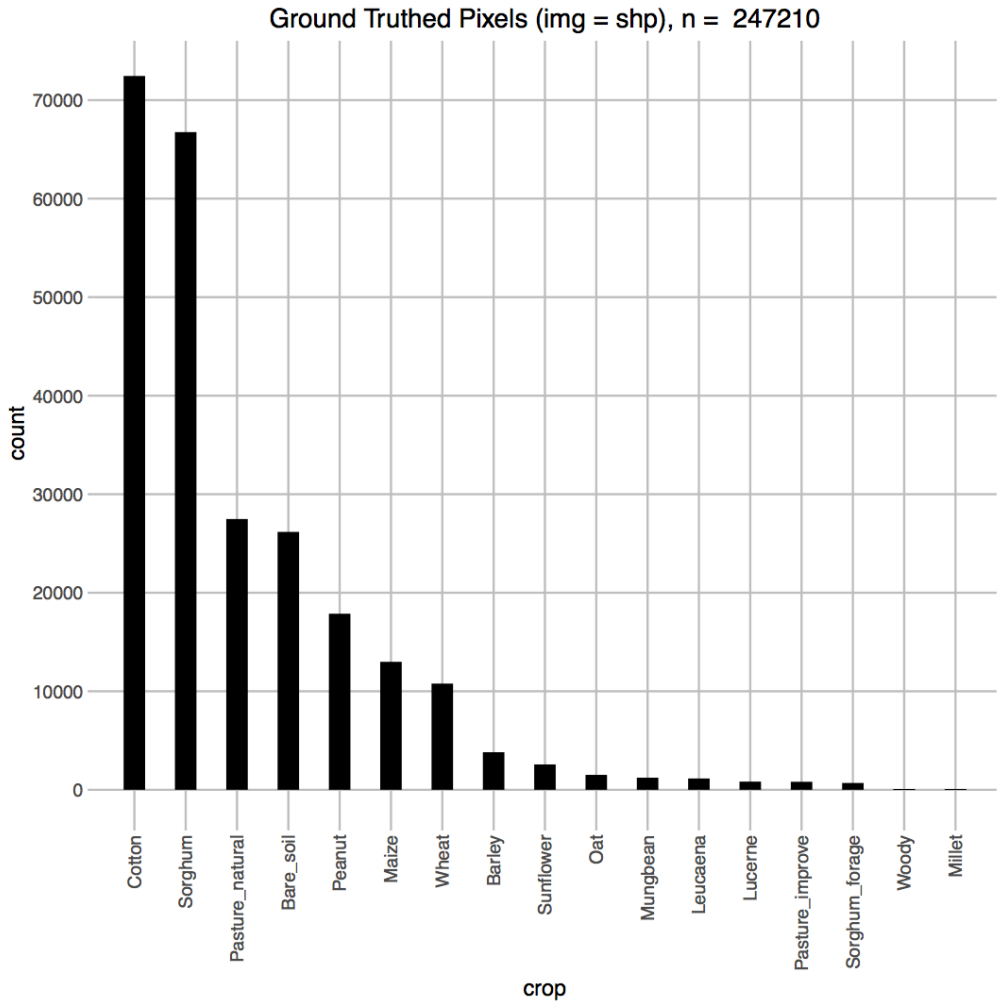
Enhanced Vegetation Index (EVI) curves for all pixels in a field growing sorghum. Image courtesy of ABS.

Variable	Description
farm.id	Field identifier
crop	Ground truth crop type
shp.year	Ground truth year
shp.month	Ground truth month
img.year	Satellite image year
img.month	Satellite image month
img.day	Satellite image day
img.time	Satellite image time
latitude	Pixel latitude
longitude	Pixel longitude
band1	Blue reflectance
band2	Green reflectance
band3	Red reflectance
band4	Near Infrared (NIR) reflectance
band5	Short–Wave Infrared 1 (SWIR–1) reflectance
band6	Short–Wave Infrared 2 (SWIR–2) reflectance

Example Data Continued

Farm.ID	Crop	shp.year	shp.month	img.year	img.month	img.day	img.time	latitude	longitude	band1	band2	band3	band4	band5	band6
3	Sunflower	2012	2	2011	11	15	23-41-40.2345	151.42238	27.269875	806	1254	1842	3094	3578	2443
3	Sunflower	2012	2	2012	1	2	23-42-06.8890	151.42238	27.269875	761	1002	1364	2158	3160	2326
3	Sunflower	2012	2	2011	11	15	23-41-40.2345	151.42263	27.269625	792	1254	1829	3128	3597	2443
3	Sunflower	2012	2	2012	1	2	23-42-06.8890	151.42263	27.269625	761	1018	1350	2192	3179	2306
43	Wheat	2012	2	2011	11	15	23-41-40.2345	151.91038	27.717125	521	737	948	1771	2345	1490
43	Wheat	2012	2	2012	1	2	23-42-06.8890	151.91038	27.717125	601	694	891	1804	2256	1385
43	Wheat	2012	2	2011	11	15	23-41-40.2345	151.91063	27.716875	551	737	945	1737	1914	1544
43	Wheat	2012	2	2012	1	2	23-42-06.8890	151.91063	27.716875	504	690	811	1867	1871	1476

Example Data Continued



Number of ground truthed pixels in each land cover category.

Example Analysis I

- Supervised Classification via artificial neural network
- Predictors: 6 spectral bands, month, year
- Vegetation indices were tried, no improvement found.
- 80/20 training/test split
- 3 layers (200 nodes layer 1, 100 nodes layer 2, 15 nodes layer 3)

Example Results I

- Overall misclassification rate 26%
- Misclassification by crop type:

Crop Type	Misclassification Rate
Bare Soil	0.399
Barley	0.507
Cotton	0.174
Leucaena	0.594
Lucerne	0.889
Maize	0.521
Millet	1
Mungbean	0.442
Oat	0.581
Pasture Improve	0.470
Pasture Natural	0.077
Peanut	0.087
Sorghum	0.286
Sorghum Forage	0.828
Sunflower	0.574
Wheat	0.300
Woody	0.901

Example Analysis II

- Supervised Classification via ridge and LASSO regularised Multinomial Logistic Regression.
- Predictors: spectral bands [1, 3, 4, 5], several vegetation indices.
- 100–fold cross validation (to estimate shrinkage parameter).
- Linear and polynomial (up to order 4) terms considered.

Example Results II

Regularization Method	Polynomial & Interaction Terms Included	% Observations Correctly Classified	Computation Time (hrs)
LASSO	No	78.16	2.12
LASSO	Yes	78.59	27.40
Ridge	No	68.54	0.49
Ridge	Yes	74.63	2.17

Overall classification accuracy of the Ridge and LASSO regularized models and the time required to fit each model.

Ground Cover Category	Total Number of Pixels in Category	LASSO LME	LASSO LME + P4 & I2	Ridge LME	Ridge LME + P4 & I2
Cotton	72 450	92.2	93.2	94.96	93.0
Sorghum	66 751	80.3	79.5	78.23	78.9
Pasture Natural	27 479	77.6	76.4	47.66	64.5
Bare soil	26 173	91.0	89.9	83.93	89.2
Peanut	17 868	82.9	85.1	26.00	66.4
Maize	12 986	14.2	21.1	0	10.6
Wheat	10 778	10.3	11.6	0.03	6.2

Accuracy by land cover category. LME = Linear Main Effects (only). P4 & I2 = Polynomial terms to order four and Interaction terms (in addition to linear main effects terms).