A Novel Agriculture Area Detection Based on CNN

Dr. T. Arumuga Maria Devi[1], A.Pravika[2]
[1]Assistant Professor, Dept of Information Techonology
[2] Dept of Information Techonology
[1, 2] Centre for Information Technology and Engineering, Manonmaniam Sundaranar University

Abstract- *In recent years the time series of Agriculture products is varied depending upon nature, and other factors. Time series and natures are very important for the crop manufacturing. The remotely image time series used to the analysis the factor. This will be processed by CNN based the active learning method. The feature of the remotely sensing images is classified accordingly sensing images is classified according to the layers, and that will analyses the dimension performs the area which produces the crop, and other things. The result of this analysis has performed remote sensing the image classification, and the accuracy of the area which can calculate at 96.2% accuracy. This framework can then help Land Use Change (LUC) monitoring as it produced similar results compared to other methods that demands more human and financial resources to be adopted.*

Keywords- Remote images, CNN, AL, ROI, SVM.

I. INTRODUCTION

BRAZILIAN southeastern and central regions economy is based on pillars in which agriculture is one of the most important. Historically, the state of São Paulo is the major national Crop (Saccharum spp.) producer, being that large investments were made in the 2000s to increase the production aiming to attend internal and external markets demand. So, between 2001 and 2010, Crop area increased from 2.5*106 ha to 5*106 ha whereas the production increased from 207*106 tons to 361*106tons.

OVERVIEW OF PROJECT

The Crop's cycle in these regions begins in April and lasts until March of the next year, comprising the growing, maturation and harvest phases. The possible climate change, one of the most important discussions nowadays, could bring several consequences to agricultural production, leading to change its current geography, forexample.

Beyond its relevance in fuel and food industries, and although it's still a recent research theme, studies indicate that C4 plants like Crop could take advantages of climate changes, precisely of atmosphere's CO2 concentration growth. This advantage is associated with the increasing of photosynthetic efficiency (and consequently of the biomass production), increasing of non-structural carbohydrates concentration (reflecting in a more efficient production) and reduction ofstomatal
conductance (that increases the efficiency of plant water use, allowing Crop to keep the production with lower water consumption).

FRAMEWORK

Proposed a framework that uses AL to optimize the training set based on class labels transferred from a domain (source) to n target images, where authors could reach up to 90% overall accuracy on image classification. Débonnaire et al. tested AL performance over traditional supervised learning methods on time series classification by different training set building heuristics. Results showed that the performance of each algorithm and the heuristic adopted depends on dataset's characteristics. In this paper, we proposed a Crop cultivated area detection framework based on time series classification so that it presents scalability and efficiency by restricting human intervention to training set building phase and using algorithims to automate classifying process in order to reduce its operational costs.

DATA MINING TECHNIQUES

Data Mining techniques are mainly divided in two groups, classification and clustering techniques. Classification techniques are designed for classifying unknown samples using information provided by a set of classified samples. This set is usually referred to as a training set as it is used to train the classification technique how to perform its classification. Generally, Neural Networks and Support Vector Machines, these two classification techniques learn from training set how to classify unknown samples.

In large data sets, data mining is the computational process for discovering new patterns. Data mining provides major advantage in agriculture for disease detection, problem prediction and for optimizing the pesticides. In recent technologies agriculture related activities provide lotof information. Hence this data mining techniques in agriculture is used for pattern reorganization and disease detection.

Data's of agriculture in data mining can be presented in form of data marts. Crop production for reliable and timely requirement for various decisions for marketing, pricing, storage distribution and import- export. The yield of agriculture primarily depends on diseases, pests, climatic conditions, planning of different crops for the harvest productivity are the results. So by these predictions are very useful for agriculture domains. Data mining techniques are used for pre-harvest forecasting. For example by applying data mining technique government can fully benefit data about farmers buying patterns and also to gain a superior understanding of their land to protect them in order to gain more profit on farmer'spart.

Data mining is also called as knowledge discovery database (KDD). Data mining tasks can be classified into two categories:

- Descriptive datamining.
- Predictive datamining.

Descriptive data mining tasks characterize the general properties of the data in the database while predictive data mining is used to predict the direct values based on patterns determined from known results. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest. As far as data mining technique is concern, in the most of cases predictive data mining approach is used. Predictive data mining technique is used to predict future crop, weather forecasting, pesticides and fertilizers to be used, revenue to be generated and soon.

Another classification technique, K- Nearest Neighbor, does not have any learning phase, because it uses the training set every time a classification must be performed. A training set is known, and it is used to classify samples of unknown classification. The basic assumption in the KNearest Neighbor algorithm is that similar samples should have similar classification. The parameter K shows the number of similar known samples used for assigning a classification to an unknown sample. The K-Nearest Neighbor uses the information in the training set, but it does not extract any rule for classifying the other.

In the event a training set not available, there is no previous knowledge about the data to classify. In this case, clustering techniques can be used to split a set of unknown samples into clusters. One of the most used clustering techniques is the K-Means algorithm. Given a set of data with unknown Classification, the aim is to find a partition of the set in which similar data are grouped in the same cluster. The parameter K plays an important role as it specifies the number of clusters in which the data must bepartitioned.

The idea behind the K-Means algorithm is, given a certain partition of the data in K clusters, the centers of the clusters can be computed as the means of all samples belonging to clusters. The center of the cluster can be considered as the representative of the cluster, because the center is quite close to all samples in the cluster, and therefore it is similar to all ofthem.

There are some disadvantages in using K-Means method. One of the disadvantages could be the choice of the parameter K. Another issue that needs attention is the computational cost of the algorithm. There are other Data Mining techniques statistical based techniques, such as Principle Component Analysis (CNN), Regression Model and BiclusteringTechniques have some applications in agriculture or agricultural - relatedfields.

1.4.1 Applications of Data Mining Techniques in Agriculture

There are number of studies which have been carried out on the application of data mining techniques for agricultural data sets. Naïve Bayes data mining technique is used to classify soils that analyze large soil profile experimental datasets. Decision tree algorithm in data minig is used for predicting soil fertility. By using clustering techniques based on Partitioning Algorithms and Hierarchical Algorithm the land utilization for agriculture and non-agriculture areas for the past ten years have been determined.

There are several applications of Data Mining techniques in the field of agriculture. Some of the data mining techniques are related to weather conditions and forecasts. For example, the K-Means algorithm is used to perform forecast of the pollution in the atmosphere, the K Nearest Neighbor (KNN) is applied for simulating daily precipitations and other weather variables, and different possible changes of the weather scenarios are analyzed using SVMs.

Supervised bi clustering technique to a dataset of wine fermentations with the aim of selecting and discovering the features that are responsible for the problematic fermentations and also exploit the selected features for predicting the quality of new fermentations. Taste sensors are used to obtain data from the fermentation process to be classified using ANNs. Similarly, sensors are used to smell milk that is classified using SVMs.

TIME SERIES

A time series is a sequence of points measured successively in time. The most obvious example for a time series is probably the Dow Jones or the development of certain stock prices. An increasingly large part of worlds data is in the form of time series [Maimon and Rokach, 2005]. But not only the economy and financial sector produce a large amount of such data.

Social media platforms and messaging services record up to a billion of daily interactions [Piro, 2009] which can be treated as time series. Besides the high dimensions of this data, the medical and biological sectorprovide a great variety oftime series, as gene expressiondata, electrocardiograms, growth development charts and manymore.

1.5.1 Time Series Data Mining

We have seen that time series can have a variety of sources as weekly sales and stock prices in the financial sector, daily temperature, earth movements, development of organisms in fields such as meteorology, seismology or phenomics, just to name a few.

The analysis objectives can be as diverse as the data sources and formats. Time series as seismograms or the Dow Jones are mostly analysed with the goal to predict its evolution for the next day's based on previous observations, in order to forecast earthquakes or unprofitable economic trends. But they can also be investigated to extract interesting or surprising trends in order to explore and understand their cause

SYSTEM OVERVIEW

Data Mining is a non-trival process of classifying the valid, novel, potentially useful and ultimately understandable patterns in data. Data Mining is used to discover knowledge from data. In other words Knowledge mining from data, knowledge extraction, data/patterns analysis, data archeology and data dredging. Multiple time series data are very perceptive to analysis and predict the disease. In multiple time series data contains multiple measurement data are collected from different time interval .The selection of the data individuality are the one of the most important drawback of dataprocessing.

There are two types of data: time series data and cross section data. Time-series data are a progression of examine of a particular feature, which are ordered in time, and cross-section data are accumulated by examining the many features at the same time. In the time series, the features can vary over time and these modifications contain important information.

Changing the data from a low level quantitative form to a high level qualitative description is called the temporal abstraction. The method of temporal abstraction takes any raw or preprocessed data as input. Temporal classification of time-related data has assured properties that discriminate it from other classification methods and by using the characteristics of temporal data in theory, there is an improvement in the
temporal classification. There are various techniques are reviewed: Support vector clustering Support vector data description, Hidden Markov Model, Pattern-Based Decision Tree, Pattern Based class-Association Rule, Classify- by-Sequence, Minimal Predictive temporal patterns

EXISTING SYSTEM

Beyond its relevance in fuel and food industries, and although it's still a recent research theme, studies indicate that C4 plants like Crop could take advantages of climate changes, precisely of atmosphere's CO2 concentrationrowth.

This is associated with the increasing of photosynthetic efficiency (and consequently of the biomass production), increasing of non- structural carbohydrates concentration and reduction of stomatal conductance.

To detect that farm field various data acquisition and detecting techniques were presented as discussed in literature like EMR color detection CO2 sensor based detection and various classification techniques for distance finding on vectors extracted from the dataset.

PROPOSED SYSTEM

The CNN approach consisted of selecting seasonal time series information from less than 1% of each class' pixels to build the training set and evaluate this selection by an expert user supported by distance measurements, repeating this process until both distance measurement thresholds were satisfied.

In most years, the classification results presented about 90% of correlation with official estimates based on both traditional and satellite image analysismethods and provide crop details.

This framework can then help Land Use Change (LUC) monitoring as it produced similar results compared to other methods that demands more human and financial resources to be adopted.

Benefits of Time Series Analysis

Parameter	Value
Data product	MOD13Q1 – NDVI; Version 5
Spatial resolution	Width: 250m Height: 250m
Sinusoidal tile grid	From h12, v10 to h14, v11
Spacial origin point	Longitude: -53.1101 degree, Latitude: -19.7797 degree
Image size	Width: 3580 pixels Height: 2214 pixels
Pixel size	0.0025 degree
Temporal resolution	16-day
Number of images	23 images/year; 230 images in total
Data acquisition period	From April, 2004 to March, 2014
Missing images	0
Coordinate system	WGS-84, geographic
Data value type	16-bit integer
Geometric correction	Already corrected
Atmospheric correction	Already corrected

Time series analysis has various benefits for the data analyst. From cleaning data to understanding it and helping to forecast future data points this is all achieved through the application of various time series models, which we'll touch on later.

- **Cleaning Data:** The first benefit of time series analysis is that it can help to clean data. This makes it possible to find the true "signal" in a data set, by filtering out the noise. This can mean removing outliers, or applying various averages so as to gain an overall perspective of the meaning of thedata. Of course, cleaning data is a prominent part of almost any kind of data analysis. The true benefit of time series analysis is that it is accomplished with little extra effort.
- **Understanding Data:** Another benefit of time series analysis is that it can help an analyst to better understand a data set. This is because of the models used in time series analysis help to interpret the true meaning of the data, as touched on previously.
- **Forecasting Data:** Last but not least, a major benefit of time series analysis is that it can be the basis to forecast data. This is because time series analysis by its very nature uncovers patterns in data, which can then be used to predict future data points.

LAND USE CHANGE

WORK FLOW OF PROPOSED SYSTEM

This study was divided into steps, as illustrated in Fig 4.1. First, time series were organized and divided into two sets representing the same area and period: one submitted through seasonal modeling and another one used to test its effectiveness. Then, both seasonally modeled (SM) and raw (R) time series were submitted through dimensionality reduction to reduce its size and eliminate remaining noise that could harm final classification.

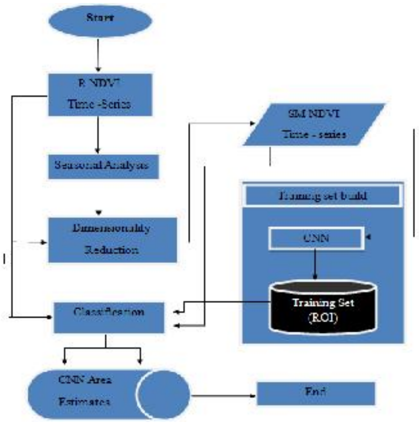


Fig. 4.1 Workflow of Proposed System

Using the CNN approach, we built the training set by sampling regions of interest (ROIs), or pixels, from SM time series images and used it to train classifiers for both SM and R time series. Finally, we compared Crop area estimates for every harvest cycle considered in this study with the analogous INPE/CANASAT project and other official estimates.

STUDY AREA AND DATA ACQUISITION

We have chosen the state of São Paulo (Fig. 4.2) due to its leadership in Crop production sector as the national main producer. The study period comprises all Crop harvests occurred between 2006 and 2012. In this period, boosted by favorable economic conjunctures, large investments were made in Crop production sector, what resulted in a great Crop area expansion dynamic.

Fig. 4.2 Sao Paulo State and its Localization in Brazil

MODIS product related to vegetation phenological behavior. This dataset was timely organized and synchronized with Brazilian center- south region Crop harvest cycle, which starts in April and lasts until March of the next year. Table 4.1 shows the dataset main chacacteristics.

Table 4.1 Input Dataset Information

Parameter	Value
Data product	MOD13Q1 – NDVI; Version 5
Spatial resolution	Width: 250m Height: 250m
Sinusoidal tile grid	From h12, v10 to h14, v11
Spacial origin point	Longitude: -53.1101 degree, Latitude: -19.7797 degree
Image size	Width: 3580 pixels Height: 2214 pixels
Pixel size	0.0025 degree
Temporal resolution	16-day
Number of images	23 images/year; 230 images in total
Data acquisition period	From April, 2004 to March, 2014
Missing images	0
Coordinate system	WGS-84, geographic
Data value type	16-bit integer
Geometric correction	Already corrected
Atmospheric correction	Already corrected

TIME SERIES ORGANIZATION AND PRE-PROCESSING

The data pre-processing is a key point in data mining projects,being responsible for data cleaning and transforming in order mining tasks results.

Here, we applied two major procedures:to improve

(i) Noise detection and removal, followed by seasonal extraction on R time series to generate SM time series, and
(ii) Dimensionality reduction on both R and SM time series.

Initially, NDVI time series were composed of 23 images/year (Table 4.1), however, Timesat tool, used in step (i), works with full seasons,
i.e. the center-most seasons. As Crop phenological behavior peaks in summer (December to March, in south hemisphere), asingleyearly time series wouldn't allow Timesat to correctly detect its seasonality. Due to this, we reorganized time series by joining 4 of them in a windowed formand
used the two center-most ones as the target for seasonal extraction (Table 4.2).

Table 4.2 Time Series Reorganization for Seasonal Modeling

Sugarcane cycle	Joined time series	Target seasons
1	2004 2005 2006 2007	2005 2006
2	2005 2006 2007 2008	2006 2007
3	2006 2007 2008 2009	2007 2008
4	2007 2008 2009 2010	2008 2009
5	2008 2009 2010 2011	2009 2010
6	2009 2010 2011 2012	2010 2011
7	2010 2011 2012 2013	2011 2012

The reason for choosing the two center-most seasons is related to Crop harvest period extension. As Crop cycle lasts at least a year, seasonal extraction must be able to detect seasons that starts both in the beginning, e.g. April, and in the end, e.g. November, of a year. Therefore, for each image pixel, time series length increased from 23NDVI values to 92 Also, its important to emphasize that cycles are not dependent on each other, since we conducted cycle analysis individually.

Noisy data can persist even after choosing the best image of each 16-day interval, and due to that, we applied, also on Timesat, the Seasonal Trend Decomposition (STL) method in order to remove spikes and outliers from time series that could harm seasonal modelling. This method is based on locally weighted regression smoother (LOESS) and produces a reminder (or residual) parameter used to weight time series values and classify then as outliers or not.
a value y(t) is classified as an outlier in two situations:

y(t)=<(Nean(y(t+1)e)……(Eqn.4.1)
y(t)=>(Nax(y(t1),y(t+1))c)(1)…..(Eqn.4.2)

where t refers to time; and c is the product of time series standard deviation and a stiffness factor (Table 4.3).

Table 4.3 Timesat Processing Parameters

Function	Parameter	Value
STL	Stiffness	3
STL	Spyke method	STL replace
DL	Amplitude	0
DL	Seasonal parameter	1
DL	Envelope iterations	3
DL	Adaptation strength	5
DL	Season start/end	Seasonal amplitude
DL	Season start	10% of amplitude
DL	Season end	10% of amplitude

Yet in step (i), the seasonal extraction was applied to model vegetation phenological phenomena in response to its phases/seasons. The double-logistic (DL) function adopted here provides both data smoothing and seasonal parameters extraction from timeseries.

Its general form is defined as:

g(t; x1, … , x4) = 1 (1 + e (x1 t)/x2⁄)1 (1 + e (x3 t)/x4⁄) ……...……(Eqn.4.3)

where, t refers to time; x1 is the inflection point in the left side of the resulting curves pike; x2 is the changing rate in x1; x3 is the inflection point in the right side of the resulting curve spike; and x4 is the changing rate in x3. We adjusted DL function hyper parameters in order to place the upper envelope of original NDVI time series curve.

This its curve at procedure is important because most noise in remote sensed data is negatively biased. Table 4.3 shows parameters values defined for STL and DL functions application on Timesat.

Several seasonal parameters related to vegetation phenological dynamics are available on Timesat. Table 4.4 shows extracted parameters from NDVI time series. Other available parameters include integrated values and were not considered for this study.

Table 4.4 Seasonal Parameters Extracted

Seasonal parameter	Description
Beginning	Season start
Ending	Season end
Length	Season lengh
Mid-season	Time to reach maximum NDVI value
Maximum	Maximum NDVI value in season
Base value	Average between minimum NDVI values
Amplitude	Between base and maximum values
Growth rate	Growth rate from season start to mid-season
Decline rate	Decline rate from mid-season to season end

Finally, in step (ii), we applied the Principal Components Analysis (CNN) on both SM and R time series using ENVI 5.2 system. This method reduces the number of attributes of the original dataset in response to the variance contained in it. The original dataset attributes were then converted into eigenvectors with eigen values that represents their variance. We can retained only eigenvectors with eigen values ≥ 0.7, what corresponds to approximately 90% of the variance contained in original datasets.

CNN – TRAINING SET BUILDING

To build the training set, we applied the AL approach by sampling, for each class, with ENVI 5.2 system,a balanced number of samples (ROIs) from SM time series images and evaluating their separability, i.e. how distinct is one class from the others, repeating this procedure until the optimal result (training set) was achieved (Fig.4.3).

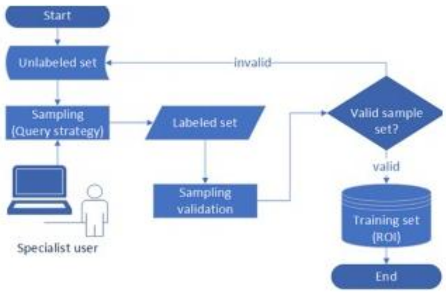


Fig. 4.3 CNN Learning – Training Set Building

CLASSIFICATION

We have used the SVM algorithm to classify the SM and R time series because it usually has good performance in remote sensing applications and can generate satisfying results even with a compact training set.

Originally, SVM is a binary classifier that defines a hyperplane to separate instances into two given classes, w0 + w1x1 + w2x2 where each instance is classified by its xi attributes weighted by wi factor. Although defines the hyperplane, it does not give optimal class separation yet. So, to define it, SVM takes some key instances, also called support vectors, and use then to define the class separation optimalmargins,

H1 : w0 + w1x1 + w2x2≥+1……(Eqn.4.4)

H2 : w0 + w1x1 + w2x2 ≤ 1 ……...……(Eqn.4.5)

where, Hi represents the optimal separation margins for classes +1 and-1.

This support vector strategy explains SVM's robustness even when the training set is small and justifies its use in remote sensing applications. Sometimes, classification does not have a linear solution. In this case, SVM uses an internal algorithm, also called kernel, to change data dimension and transform a non-linear separation problem into linear. In this paper, we used the Radial Basis Function (RBF), considered one of the best kernels available in literature as:

(,Xj)=e‖XiXj‖2/2a2…..……...……(Eqn.4.6)

where, xi and xj are instance attributes. We conducted the SVM hyper parameters configuration adopting the default values for most of then. Particularly, for Classification probability threshold (Table 4.5), we empirically defined its value as 0.5, i.e . the minimum probability required to SVM classify a pixel (a time series) is 50%. Table 4.5 shows hyper parameters configuration for SVM used in this study

Table 4.5 SVM Hyper parameter List

Hyperparameter	Value	Value description
Kernel type	RBF	Presented in [54]
Gamma in kernel func. (γ)	$f(x) = 1/x$	x = no. of bands.
Penalty parameter (C)	100.00	Default value
Pyramid levels	0.00	Default value
Classif. prob. threshold	0.5	Empirically defined

Configuration avalaible on ENVI 5.2 system. Gamma in kernel function value is the software default.

VALIDATION

After classification, all classes, excepting Crop, were merged into a single class (no-Crop) to validate model's performance. Correspondent Crop area estimates from INPE/CANASAT project were used as the test set (ground truth) to validade SVM classification for both SM and R time series. In this step, as opposed to, we performed a totally random sample from the test, set aiming to avoid any bias in model's validation, e.g. intentionally selecting regions with higher or lower Crop area concentration.

The test set size was dimensioned to have, as usually employed, 20% of the training set size. We also compared both SM and R time series area estimates with official government agencies like National Food Supply Company (acronym CONAB, in Portuguese), IBGE and INPE/CANASAT itself.

SOFTWARE DESCRIPTION

ASP.NET FRAMEWORK

Front-end Environment (.NET Framework)

The Internet revolution of the late 1990s represented a dramatic shift in the way individuals and organizations communicate with each other. Traditional applications, such as word processors and accounting packages, are modeled as stand-alone applications: they offer users the capability to perform tasks using data stored on the system the application resides and executes on.

Most new software, in contrast, is modeled based on a distributed computing model where applications collaborate to provide services and expose functionality to each other. As a result, the primary role of most new software is changing into supporting information exchange (through Web servers and browsers), collaboration (through e-mail and instant messaging), and individual expression (through Web logs, also known as Blogs, and e- zines - Web based magazines). Essentially, the basic role of software is changing from providing discrete functionality to providing services.

The .NET Framework represents a unified, object-oriented set of services and libraries that embrace the changing role of new network-centric and network- aware software. In fact, the .NET Framework is the first platform designed from the ground up with the Internet in mind.

Microsoft .NET Framework is a software component that is a part of several Microsoft Windows operating systems. It has a large library of pre- coded solutions to common programming problems and manages the execution of programs written specifically for the framework. The .NET Framework is a key Microsoft offering and is intended to be used by most new applications created for the Windows platform.

Benefits of the .NET Framework

The .NET Framework offers a number of benefits to developers:

- A consistent programming model
- Direct support for security
- Simplified development efforts
- Easy application deployment and maintenance

.NET CLASS LIBRARY

The .NET Class Library is a key component of the .NET Framework it is sometimes referred to as the Base Class Library (BCL). The.NET Class Library contains hundreds of classes you can use for tasks such as the following:

- Processing XML
- Working with data from multiple data sources
- Debugging your code and working with event logs
- Working with data streams and files
- Managing the run-time environment
- Developing Web services, components, and standard Windows applications
- Working with application security
- Working with directory services

The functionality that the .NET Class Library provides is available to all .NET languages, resulting in a consistent object model regardless of the programming language developer's use. Components of the
.NET Framework

- Common Language Runtime
- .NET Class Library
- Unifying components

Microsoft Intermediate Language often referred to as IL. When your code executes for the first time, the LR invokes a special compiler called a Just In Time (JIT) compiler, Because all .NET languages have the same compiled representation, they all have similar performance characteristics. This means that a program written in Visual Basic .NET can perform as well as the same program written in Visual C++.NET.

C# is a general-purpose, modern and object-oriented programming language pronounced as "C sharp". It was developed by Microsoft led by Anders Hejlsberg and his team within the .Net initiative and was approved by European Computer Manufacturers Association (ECMA) and International Standards Organization (ISO). C# is among the languages for Common Language Infrastructure and the current version of C# is version 7.2. C# is a lot similar to Java syntactically and is easy for the users who have knowledge of C, C++ or Java.

.Net applications are multi-platform applications and framework can be used from languages like C++, C#, Visual Basic, COBOL etc. It is designed in a manner so that other languages can use it.

.NET FRAMEWORK ADVANTAGES

- ***Easy to start:*** C# is a high level language so it is closer to other popular programming languages like C, C++, and Java and thus becomes easy to learn for anyone.
- ***Widely used for developing Desktop and Web Application:*** C# is widely used for developing web applications and Desktop applications. It is one of the most popular languages that is used in professional desktop. If anyone want to create Microsoft apps, C# is the go to language.
- ***Community:*** The larger the community the better it is as new tools and software's will be developing to make it better. C# has a large community so the developments are done to make it exist in system and not become extinct.
- ***Game Development:*** C# is widely used in game development and will continue to dominate. C# integrates with Microsoft and thus has a large target audience. The C# features such as Automatic Garbage Collection, interfaces, object oriented etc. makes C# a popular game developing language.

MODULES

- Admin Module
- Student Module
- Company Module
- Job search module
- Placement module

Modules Description

- ***Admin Module:*** Admin should manage users and employee account in various ways if any user found that abuse of site that admin can ban him. Admin also can maintain master level database details and can fill entry in the various master table fields like city, state, country, products, categoriesetc.
- ***Student Module*:** Student is registered by this site. A student can apply job for company and eligible students give a online exam which is held bycompany.
- ***Company Module*:** Company is a person with a given or not given appropriate authentication and privileges. Company can perform all level tasks such as post of job and take an online exam as well as well as they manage our status through thesite.
- ***Job search Module:*** Most people don't particularly enjoy the process of job hunting; however, it is necessary if you are looking for work. These days, we have a lot of different resources available to assist us with our job search, making the process a lot easier. It is important to ensure you have a good CV before applying for a job and that you tailor it to suit the job you are applying for. You should also send a

cover letter with each application. These are some methods to use when searching for a jobonline.

- ***Placement Module:*** The work placement module allows the students develop skills that are critical to the workplace. In the placing of any student it was essential that care was taken to ensure that prior to any student was placed that the role was deemed appropriate to the level expected from thestudent.

RESULTS AND DISCUSSION

SEASONAL MODELING

Seasonal extraction showed that value-based parameters performed higher vegetation cover contrast than time based parameters. Temporal parameters colored compositions (Fig. 6.1) were more homogeneous when compared to value-based ones (Fig. 6.2), evidencing that these parameters are better to be used in LU studies, as it provides more information about different vegetationtypes.

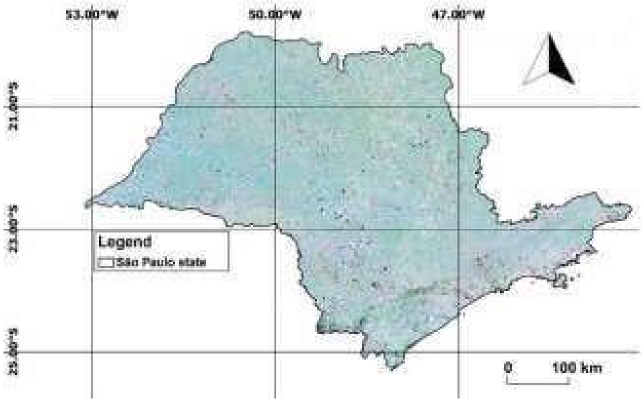


Fig. 6.1 Colored Composition of Time Based Seasonal Parameters

Although temporal composition did not show differences between vegetation types, we observed that temporal seasonal parameters agree with 1961-1990 Climate Normalsobserved for São Paulostate.

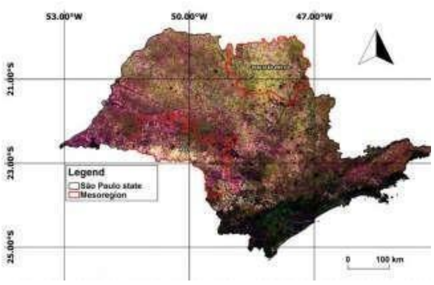


Fig. 6.2 Colored Composition of Value Based Seasonal Parameters

The homogeneity of seasonal compositions using temporal related parameters means that all vegetation types have similar temporal behavior.

OVERVIEW OF THE DATASET

The data available in this proposal is obtained for the years from 1965 to 2009 in East Godavari district of Andhra Pradesh in India. The data is taken in four input variables. They are Year, Rainfall, Area of Sowing and Production. Year attribute specifies the year in which the data available in Hectares. Rainfall attribute specifies the Rainfall in East Godavari in the specified year inCentimeters.

Area of sowing attribute specifies the total area sowed in East Godavari district in the specified year. Production attribute specifies the production of crop in East Godavari district in the specified year in Tons. The preliminary data collection is carried out for all the districts of Andhra Pradesh in India. Each area in this collection is identified by the respective longitude and latitude of the region. In this paper the evaluation is considered for only one district i.e. EastGodavari.

The information gathering process is done with three government units like Indian Meteorological Department, Statistical

Institution and Agricultural department. Instead of restricting with few regions and few samples of data, it is aimed at applying the Expectation Maximization approach on all the regions of Andhra Pradesh in India. In this paper the estimation of the crop yield is analyzed with respect tofour

DIMENSIONALITY REDUCTION

CNN analysis of SM and R time series reduced the original dataset dramatically. The 9 original attributes of SM time series were transformed into 2 attributes keeping 87% of data variance in every harvest cycle. In case of R time series, CNN reduced from 46 attributes to 2 or even 1 attribute (in case of 2006 and 2009 cycles), keeping an average data variance of 93% of the original dataset. The number of retained eigenvector and variance in this study is similar, which retained three (with 99% of the original variance) and one (with an accuracy of 90% in final classification) eigenvector, respectively.

CONVOLUTIONAL NEURAL NETWORK

A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.class depending on pairwise combination and period of year, they didn't define a fixed distance threshold.

VALIDATION

The model validation with test samples showed that classification results using SM time series (Table 6.1) were a few better than those using R time series for minority (Crop) class (Table 6.2). In general, average Kappa index for SM time series classification was 0.25 against 0.21 from R time series, what is still weak, although it is still better than aleatory classification. Also, as mentioned before, we have performed a totally random sampling on test set in our experiments

.

Table 6 Table 6.1 Validation of Seasonally Modelled Time Series or Min for Minority (Crop) Class Classification

Year	False positive rate	False negative rate	Recall	Specificity	Precision
2006	67.16%	46.84%	51.10%	80.00%	32.84%
2007	61.11%	70.21%	29.79%	90.06%	38.89%
2008	61.54%	62.26%	37.74%	85.12%	38.46%
2009	48.87%	71.88%	28.13%	91.47%	51.45%
2010	53.79%	54.41%	46.56%	84.00%	49.21%
2011	51.56%	53.03%	46.97%	83.60%	48.41%
2012	58.67%	56.34%	43.66%	77.68%	41.33%
Average	57.06%	59.57%	40.43%	84.39%	42.94%

In this case, what could had happened is that we randomly sampled test ROIs from small Crop areas, as in São Paulo state exists large continuous Crop areas as well as small spread areas.

Table 6.2 Validation Raw Time Series for Minority (Crop) classification

Year	False positive rate	False negative rate	Recall	Specificity	Precision
2006	80.00%	83.72%	16.28%	87.56%	20.00%
2007	65.57%	55.32%	44.68%	81.90%	34.43%
2008	50.00%	71.70%	28.30%	93.02%	50.00%
2009	67.42%	54.69%	45.31%	70.59%	32.58%
2010	42.86%	70.59%	29.41%	92.50%	57.14%
2011	56.99%	39.39%	60.61%	73.76%	43.01%
2012	61.86%	36.62%	63.38%	62.94%	38.14%
Average	60.57%	58.86%	41.14%	80.32%	39.33%

Also, even been considered a reliable project, remainder noise on INPE/CANASAT data, taken here as the test set, can still be present, as it involved visual image interpretation tasks which is susceptible to errors as well.

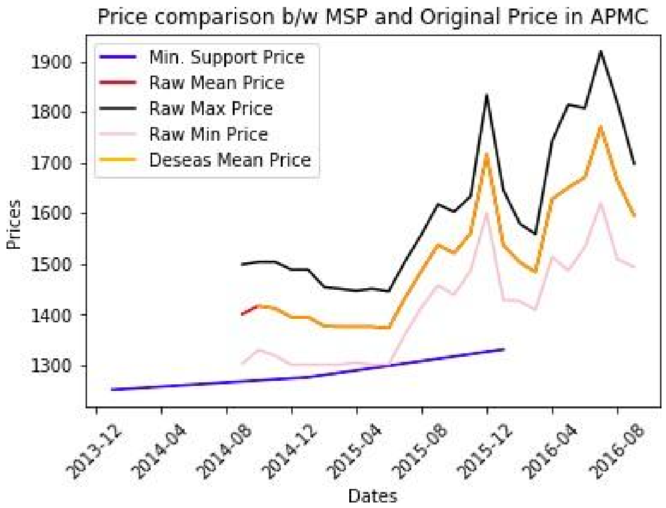
SVM applications in remote sensing focus most on hyperspectral classification, so there is a gap in time series classification using it. In recent studies, authors combine time series with other data sources in order to generate their results. In this paper, we have used only NDVI time series in our analysis. Carrão, Gonçalvesand Caetano, classifying vegetation index time series, concluded that spectral diversity is more important than temporal diversity in order to obtain satisfactory results.

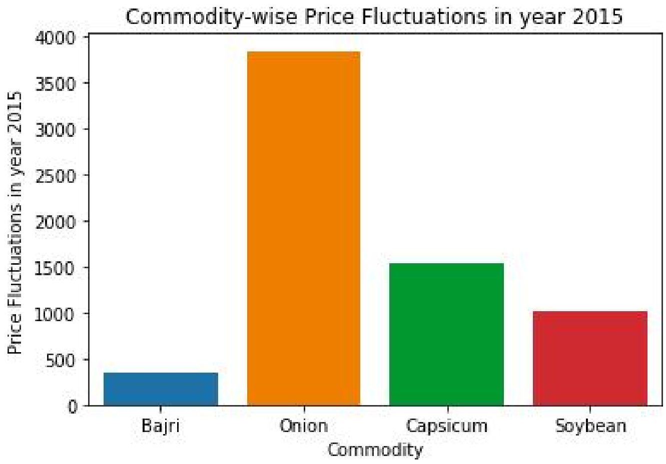
Despite of these evidences, São Paulo area has a peculiar characteristic. As it is the major national Crop producer, it still has a considerable pasture area, and this type of vegetation has similar spectral behavior to Crop, what could epresent a challenge for LU assessment.

CROP AREA

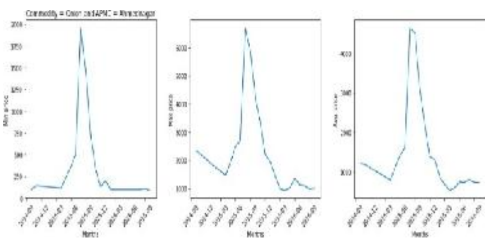
IBGE estimates, taken as official by both government and Although results observed in spatial validation results (Tables 6.1 and 6.2), when we compared the total area estimates with other estimates and when we visually compared INPE/CANASAT mapping (Fig. 6.3) with generated maps, we realized similarities between then, more on SM time series than R time series.

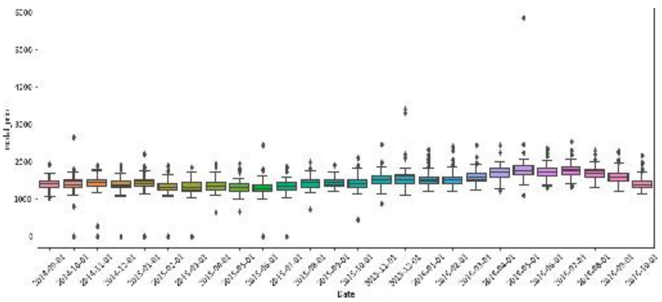
Comparing the Pearson (r) correlation for 2006 to 2012 cyles, we assessed that SM time series estimates presented 23% correlation with both INPE/CANASAT and CONAB and 24% with IBGE, whereas R time series estimates presented 59% correlation with PE/CANASAT, 63% with IBGE and a 67% with CONAB

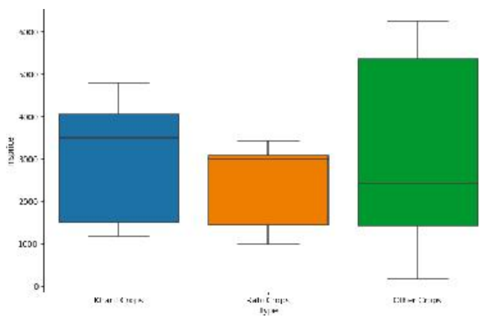




C







For 2006 and 2009 cycles, respectively, SM time series estimates differ, in average, 29% and -45% from all official estimates (39% and -46% considering only INPE/CANASAT estimates).

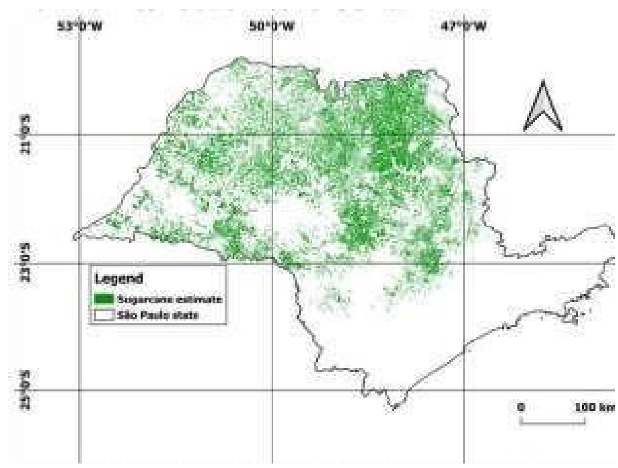


Fig. 6.3 INPE/CANASAT Mapping of Crop planted Area

For these two periods, the dimensionality reduction step coincidentally retained just one eigenvector on R time series CNN application, what suggests that some noise from input data set persisted even after applying seasonality extraction and dimensionality reduction. Discarding those cycles for both SM and R time series results, Pearson (r) correlation for SM time series estimates increased to 90% with INPE/CANASAT, 84% with CONAB and 95% with IBGE, whereas for R time series estimates, maximum correlation found is67%.

II. CONCLUSION

The time series NDVI classification by CNN depends on all the crops. And this will be predicted As the new crop area as well as the area not used in agriculture in previous has made the production. This result predicts the accuracy of the vegetable and crops with the approximate level of the production demand. The production has analyzed and the demand has been analyzing by training dataset.

The CNN Approach build efficient dataset and reduce the construction efforts. In future the demand based production of Agriculture with natural climatic calculated easily through production data set.

III. SCOPE FOR PHASE II

This way, the use of CNN in training set building can improve model's performance and efficiency by reducing its operational costs and adding technical criteria for training instances selection.

The Crop area estimates by seasonal extraction and classification generated results that are similar to other methodologies estimates. This kind of product can be used along with those methodologies in order to improve and assess model's performance.

In this proposal, used only NDVI time series in our analysis, but adopting other vegetation index (e.g. Enhanced Vegetation Index – EVI) could help to improve our results and increase model's accuracy.

REFERENCES

[1] A.P. Souza, M. Gaspar, E.A. Silva, E.C. Ulian,A.J. Waclawovsky,

[2] R.V. Santos, M.M. Teixeira, G.M. Souza, and S. Buckeridge, "Elevated CO2 increases photosynthesis, biomass and productivity, and modifies gene expression in Crop," Plant, Cell & environment, vol. 31, no. 8, pp. 1116-1127, available at https://bit.ly/2PBjysR, Aug.2008.

[3] B.F.T. Rudorff, D.A. Aguiar, W.F. Silva, L.M. Sugawara, M. Adami, and M.A. Moreira, "Studies on the rapid expansion of Crop for ethanol production in São Paulo State (Brazil) using Landsat data," Remote Sensing, vol. 2, no. 4, pp. 1057-1076, available at https://bit.ly/2L4cG4X, Apr.2010.

[4] D.B. Leakey, M. Uribelarrea, E.A. Ainsworth, S.L. Naidu, A.Rogers,

[5] D.R. Ort, and S.P. Long, "Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO2 concentration in the absence of drought," Plant Physiology, vol. 140, no. 2, pp. 779- 790, avaliable at https://bit.ly/2oB7hJi, Feb. 2006.

[6] F.R. Marin and D.S.P. Nassif, "Mudançasclimáticas e a cana- deaçúcar no Brasil: Fisiologia, conjuntura e cenáriofuturo," RevistaBrasileira de EngenhariaAgrícola e Ambiental, vol. 17, no. 2, pp. 232- 239, avaliable at https://bit.ly/2PZlD2J, Jan.2013.

[7] G.M.S. Câmara, and E.A.M. Oliveira, Produção de cana-de-açúcar. Piracicaba, São Paulo: FEALQ, pp. 1-242,1993.

[8] G.W.Wall,T.J.Brooks,N.R.Adam,A.B.Cousins,B.A.Kimball,

[9] P.J. Pinter Jr., R.L. Lamorte, J. Triggs, M.J. Ottman, S.W. Leavitt, and A.D. Matthias, "Elevated atmospheric CO2 improved sorghum plant water status by ameliorating the adverse effects of drought," New Phytologist, vol. 152, no. 2, pp. 231-248, available at https://bit.ly/2CbPCS2, Nov.2001.

[10] H.S. Pinto and E.D. Assad, Aquecimento global e cenáriosfuturos da agriculturabrasileira. Campinas, São Paulo: EMBRAPA, pp. 1-82, 2008.

[11] J.A. Marengo, Mudançasclimáticasglobais e seusefeitossobre a biodiversidade: caracterização do climaatual e definição das alteraçõesclimáticas para o territóriobrasileiroaolongo doséculo

[12] XXI. Brasília, Federal District: MMA, pp. 1-212, 2007.

[13] R.R.V. Gonçalves, P.P. Coltri, A.M.H. Avila, L.A.S. Romani, J. Zullo Jr., and H.S. Pinto, "Análisecomparativa do climaatual e futuro para avaliar a expansão da cana-de-açúcarem São Paulo," Proc. 17th CongressoBrasileiro de Agrometeorologia (CBAGRO '11), pp. 1-5, Jul.2011.

[14] R.S. Lunetta, J.F. Knight, J. Ediriwickrema, J.G. Lyon, and L.D. Worthy, "Land-cover change detection using multi-temporal MODIS NDVI data," Remote Sensing of Environment, vol. 105, no. 2, pp. 142- 154,

[15] S.J. Wand, G. Midgley, M.H. Jones, and P.S. Curtis, "Responses of wild C4 and C3 grass (Poaceae) species to elevated atmospheric CO2 concentration: a meta-analytic test of current theories and perceptions," Global Change Biology, vol. 5, n. 6, pp. 723-741, avaliable at https://bit.ly/2LRUAmK, Aug.1999.

[16] União das Indústrias de Cana-de-Açúcar, "Dados da produçãocanavieira", UNICADATA,https://bit.ly/2wBqSNa.2016.