Indian Journal of Science and Technology, Vol 9(28), DOI: 10.17485/ijst/2016/v9i28/98393, July 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Approach to Analyse the Agriculture Acreage and Estimate Production

Virrat Devaser and Ashish Kr. Luhach*

School of Computer Science and Engineering, Lovely Professional University, Jalandhar-Delhi G.T. Road, National Highway 1, Phagwara – 144411, Punjab, India; virrat.14591@lpu.co.in, ashishluhach@acm.org

Abstract

Background/Objective: Crop acreage and production estimate is of great challenge and requires strategic decisions to be taken by government agencies regarding controlling import and export duties and also to control inflation as a whole. **Methods/Statistical Analysis:** A lot of techniques exist to measures the yield and production for a given area and that too crop wise. We estimate the crop acreage with the help of remote sensing data and geographic information system techniques. In remote sensing, we acquire information related to the earth surface and to determine the crop acreage with the help of remote sensing images. We extract the useful information related to the yield estimation. **Findings**: In image processing, it is used to analyze or manipulate the features of an image. We used K-mean clustering and Self Organizing Map algorithms here for the crop Acreage and production estimation as a comparative approach to measure out acreage for a specific area image. By applying different image processing techniques and using the remote sensing techniques we estimated the production estimates for crops and compared the results with actual data obtained from statistical sources. Geographic Information System is widely used in crop acreage and production estimation to find out the locations of a particular image and helps in improving the accuracy of estimates. The challenge still remained regarding the crop segregation which requires lot of ground truth validation and hence makes the approach less viable as it reduces accuracy. Self organizing maps were found to be giving better results than other approaches. **Applications/Improvement:** It can be used by government agencies for taking decisions regarding import/export duty control and by agencies interested in procurement of crops and can be used by RBI to tackle Inflation as well.

Keywords: Remote Sensing, Geographic Information System, Un Supervised Classification, K Means, SOM Algorithm, Supervised Classification

1. Introduction

Production estimation is used for measuring the size of land in crop acreage. There are many approaches to crop area estimation in agricultural statistics that are collected with the help of census. This is one of the cheapest methods which are used in statistical analysis in which a predetermined number of observations have been taken from a large sample. In this method from a larger sample which depend on the types of analysis being performed. This process is known as sampling. Crop area and product estimation describes a "pure pixel sampling" in image. It is used for multiple methods such as area sampling, point sampling and list sampling. In area sampling, the region is sub-divided into smaller blocks. Point sampling is also known as the nearest neighbour in image. Combining area and list frames is known as multiple frame sampling. It provides good qualities and solves some of the problems. In the word 'estimate' will be used interchangeably for estimates or forecasts. With the advancement in technology these days and abundant availability of satellite images GIS (Geographical information systems) techniques have become popular coupled with Image processing techniques. The overall accuracy of estimation has increased with the use of these technologies.

2. Background and Terminology

There are many approaches to crop area estimation. Agricultural statistics are collected by censuses, which

*Author for correspondence

require numeration of the total population of interest. Sample requires enumeration of a small part of the population through Geographic information system and remote sensing. It is designed to capture, store and manage the geographic data. It is used to find out the locations in the earth space time may be recorded, in a graph. Remote sensing is used for acquiring information related to the earth surface and atmosphere using a sensor. It involves the different activities of observing the objects. The important part of remote sensing is electromagnetic radiation. Remote sensing images represent a form of digital images. It is used to extract information for image processing technique. Remote sensing is used to acquire information about the land, earth, and atmosphere using sensors or satellite platforms. Its estimates have progressed from sums of local area information, direct expansions of statistically sampled data, classification of crop specific pixel. Small area estimation is much more accurate and timely as compared to large area estimation.

Remote Sensing: Remote sensing is the activity of recording, observing or sensing the objects at remote places. Sensors are not directly interconnect with the objects has been observed. Electromagnetic radiation is mostly used in remote sensing. The output is generally seen as the image has been observed. Image analysis is required to acquire the useful information of an image. Example of remote sensing is human visual system. To acquire information about the earth surface, atmosphere using sensors or satellite, space platforms.

Satellite Remote Sensing: These are equipped with sensors looking down to the earth. They observing the earth they go round in predictable orbits.

Effects of Atmosphere: As Sensors are looking with a layer of atmosphere separate the sensors from the earth surface has been observed and the effect of atmosphere on the electromagnetic radiation travel from the earth to the sensor with the help of atmosphere. It constitutes cause wavelength dependent to absorb and scattering of radiations. These effects degrade the quality of images. Atmospheric effects can be corrected with the help of further analysis and interpretation of an image. Figure 1.

Wavelength band in the electromagnetic spectrum is strongly absorbed by the atmosphere. To determine the remote sensing by the penetrate atmosphere. The wavelength regions are known as the atmospheric Transmission windows. Remote sensing is designed to operate with one or more of the atmospheric transmission windows. These windows apply in the microwave region, wavelength band, visible region and part of the ultraviolet region. In Optical and Infrared Remote Sensing, the optical sensors detect solar radiation is effected for the earth. The wavelength region is converted from the visible and near infrared to the short wave infrared. Various materials such as water bodies, soil, vegetation covers and buildings reflect visible and infrared light. Interpretation requires the knowledge of the spectral reflectance and their signature over the surface of the earth.

Microwave Remote Sensing: They carry active or passive microwave sensors. These sensors emit electromagnetic pulses of microwave radiation to clarify the area of an image. Advantage of microwave remote sensing is they can penetrate clouds. High resolution images of the earth are generated through the Synthetic Aperture Radar. The image quality depends on the amount of waves scattered back by the target and received by the SAR antenna. It is responsible for backscatter is different from microwave compared to the radiation, interpretation of SAR image require the knowledge of how microwave interact with targets.

Remote Sensing Images: It is represented in the form of digital images. It is used to extract the useful information from the image. It may be employed to improve the image with the help of visual interpretation, to correct or restore the image which has been blurred or degraded. To improve the quality of an image with the help of some clustering algorithms one can use Hierarchical clustering, K-means clustering, Gaussian mixture models or Self-organizing maps. Figure 2.

Hierarchical Clustering: It is a method to analyse the cluster of an image. It is also known as hierarchical cluster analysis in data mining. Data is not divided into a particular cluster in a single step. Items are arranged

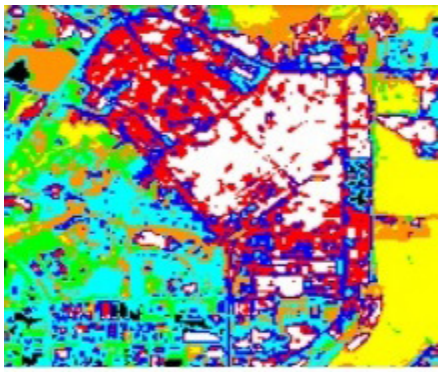

Figure 1. Demonstrating the different spectral signatures differentials.

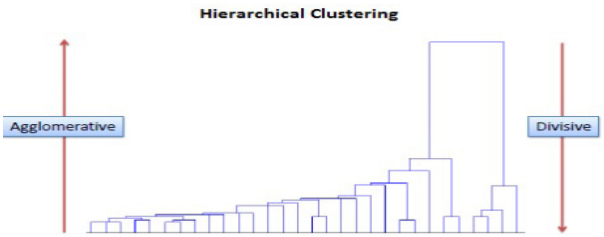

Figure 2. Demonstrating the Hierarchical Clustering Approach.

in hierarchy with a tree like structure. It consists of two types- Agglomerative and Divisive. Agglomerative is based upon the bottom up approach. It begins from bottom to top. It is used to measure the distance between nearest data point or distance of centroid whereas in case of Divisive it is a top down approach. It starts from top to bottom.

K-means clustering: K–mean cluster is used to choose the one data item in a cluster. These clusters are non-hierarchical in nature. They do not overlap with each other. The data is partitioned into the k-mean clusters. It calculates the distance between two clusters. It is a method of quantization based from processing the signal that is used for cluster analysis through data mining. It is easy to implement and use on large sets of data. It has been mostly used in agriculture, segmentation, and computer vision.

Gaussian Mixture Model: To acquire a data in between the sub part into the overall part of a data. It is mostly used in biometric systems. It consists of a mixture of multivariate components. Each component is defined by its mean, co-variance.

Self-Organizing Map: It is a data visualization method which is used to reduce dimensions of a data with the help of self-organizing neural network. It is helpful to create the high dimensional data. It provides for reduce the dimensions and display the similarities. It has two components sample data and weights. It is actually a type of artificial neural network which is trained by unsupervised learning to produce a low dimensional, discrete representation of the input space of the training samples called a map. It is different from artificial neural network because in SOM we use a neighbourhood function to preserve the topological properties of the input space. SOM is useful for visualizing low dimensional views of high dimensional data into the multi-dimensional scaling. A component of SOM is nodes or neurons. It describes a mapping from a higher dimensional input space to lower dimensional input space. Advantages of self-organizing map are that they are easy to understand easy to find out the quality of an image. Whereas disadvantages of self-organizing map are that it is very difficult to estimate the missing data of an image.

[1]Discussed about challenges will be occurred to precise the systematic classification principles and to describe the distribution of soil material and three dimensional layers. They compared hierarchical clustering with k – mean classification to create soil layer models. The classification method was used to identify the soil layer attributes. The model described the spatial variables in soil layer material among a landscape. Crop growth which is dependent on soil characteristics like that texture, bulk density or penetration resistance. A Three dimensional soil layer models were used which was beneficial for farmers to manage the soil or land. [2]Worked on finding errors associated with yields and improve the forecasting process based upon variables which was more based upon statistical way to forecast the yields. The authors of[3] worked on methodology supported by remote sensing and sampling methods making it efficient to control the accuracy of estimation. The paper[4] was based upon spatial and temporal resolutions and was used for vegetation and moderate resolution imaging spectro-radiometer. The authors of[5] used the field conditions data in addition to remote sensing data using remote sensing for data normalized were commonly used for to estimate the time period for rainfall, weather forecast, evaluation of vegetable growth or productivity. NDVI is used for various methods from simple image to more complicated image. Methods can be used for to select a crop region in Iowa represents the form of a rectangle. Data was categorized as NDVI, Precipitation, and temperature, soil moisture. The authors[6] used Crop acreage estimation for crop production estimates using remote sensing. It was done to classify the images into a number of clusters with mean of self - organizing map, artificial neural network method. SOM is a kohonen neural network is unsupervised learning models that classify the units in the same class. The authors of[7] presented the study of rice recording and estimate the product using software for estimation of yields for rice crop. The researchers[8] used two neural networks for applied self - organizing map and multilayer perception neural networks. It was based upon to classify the images at pixel level and sub-pixel level. The performance of the Self-organizing map was better estimation surface than the multilayer perception.

The authors of[9] worked on identifying the two main issues for crop production estimation. The issues were crop identification and acreage estimation. In object oriented classification was found good for crop identification. Key segmentation was used in object oriented classification.

The work presented in[10] on monitoring the Land quality for estimation of yield to calculate the remote sensing and single crop season. Land productivity was classified into three parts that is high, medium, low productivity. It was based on different standard zone. Methods were applied, to improve the gain production with the help of increasing a cultivated area, increasing the proportion of grain crops, improved multiple cropping index and improve crop yield estimation of yield was calculated with ground biomass and harvest index with remote sensing data. The research work carried in[11] on crop acreage and expected yield based on remote sensing and the types of sampling. Yields were analysed for soyabean crop through remote sensing data. MODIS satellite data and topographical sheets were deployed for remote sensing and GIS analysis.

In the research paper[12] crop acreage estimation was used to estimate the crop production using remote sensing and the effect of resolution based upon quantitative and qualitative view was analysed. Spatial statistics methods were used to analyse the data with different spatial resolutions. Three ways of maize area extraction were studied, to establish the rescaling series data, to analyse the accuracy of crop acreage estimation of different scales. To evaluate the accuracy of standard deviation and RMSE, the authors of[13] used Kohonen maps and were able to solve the supervised and unsupervised problems in easier way. The approach used was implemented by means of sequential algorithm or batch training algorithm. The important features of kohonen maps are the collection of function and algorithms are provide as matrix laboratory source files. Data scale variables are always range between zero and ones. In self-organizing map toolbox can be run via the graphical user interface. It can be easy to analysed, loaded results in the graphical user interface. In supervised classification model was calculated by using CP – ANNS, SKN's, and XY-FS with the Matlab command window. The research work carried out in[14] used Image processing techniques to estimate the vegetation indices, irrigation land mapping and for analysing the different parameters. It can also be used in agriculture field such as imaging technique, weed detection, fruit grading. The authors of[15-17] worked using the GIS and raster data to identify the area under crops. A mix of regression approach was used to determine the cropping plans obtained the yields for different farms.

3. Objectives

The main objective is to estimate the crop acreage and product estimation with the help of unsupervised learning and compare their results for efficient estimation of crops yields.

- Hierarchical clustering approach: It is used as a method of cluster analysis which aims to build a hierarchy of clusters.
- K-Means clustering: It is used as a method of vector quantization. It is used in cluster analysis for data mining[18]. It provides to partition and observations into N clusters, in which every observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
- Gaussian mixture models: It is used as a multivariate distribution that consists of a mixture of one or more multivariate Gaussian distribution components. Concept of mean and covariance is used to classify.
- Self-organizing maps: It is used as a type of artificial neural network that is learn using unsupervised learning to produce a low dimensional, discrete representation of the input space of the training samples, called a map. It consist of component called nodes or neurons. It is based on unsupervised, competitive learning.

4. Research Methodology

Geo-referencing: Physical space that associated with some locations refers to as geo-reference[18]. The process is associated to physical map or raster image of map that is in geographic information system. For example points of interest, buildings are the geographical locations where geo-referencing can be applied. Administrative Boundary Overlaying: It includes the list of land make use of ecological requirements. Agro ecological technique is based upon the list of land resources including climate and soils. To estimate about a land, the important application is land productivity assessment. Sub setting study area: multiple areas are divided into single ones. Classification is done to classify the data in different fields. It includes supervised, unsupervised and time series based techniques.

The area of sample image was having major rice crop to find the estimate of rice crop was analysed using image processing and the signature of pixels were found for the crop area. The signatures of fodder and other crops were differentiated with pixel value and a mix of ground verification. The major challenge was area classified as waste land and other type of vegetation. Therefore in order to find accurate Rice Crop area we extracted the signatures a factor by taking previous data and applied a proportionate factor that is proportion to the image. Figure 3.

5. Expected Outcomes

The expected yields efficiency was expected from the image so that quite in advance estimation of crop yields could be calculated for both Rabi as well as Kharif crops. As we have compared various techniques and applied them differently and measured out the optimum results as by comparing results through different approaches we were able to get smoothness in end results.

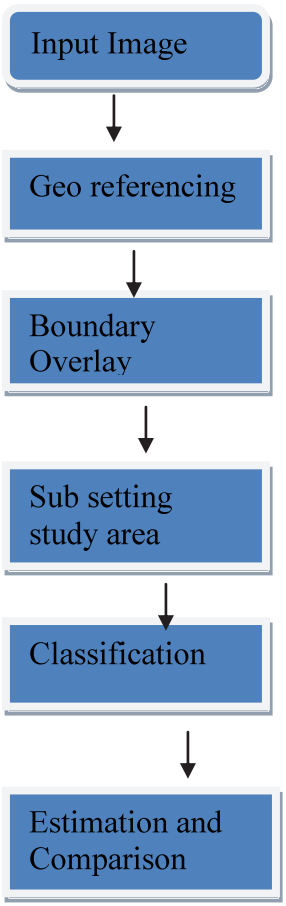

Figure 3. Summarizing the approach.

6. Conclusions

There is a lot of scope and alternatives to enhance the accuracy of the predictions made by the clustering algorithms. Agriculture resources are the important reusable natural resources. They help in making strategic decisions towards achieving food security. We can also apply similar techniques to measure out an estimate of damage done to specific crops as a result of unwanted rains or because of some other natural or artificial events that act as challenge to the health of crop. This paper uses a mix of approaches used for image processing techniques for efficient estimation of crop yields and to find an efficient measure of expected yields. The same approach can be extended to study of other crops in kharif as well as rabi seasons. The biggest challenge is the extraction of pixels of crop for non crop and that of crop from non crop. A mix of statistical and spatial data approach can be applied in future to improve upon the efficiency of the estimation.

7. References

1. Prasad AK, Chai L, Singh RP, Kafatos M. Crop yield estimation model for Iowa using remote sensing and surface parameters. International Journal of Applied Earth Observation and Geo information. 2006 Jan 31; 8(1):26-33.
2. Maurya AK. Estimation of acreage & crop production through remote sensing & GIS technique. Hyderabad, India: Proceedings of Geospatial World Forum. 2011; p.1-14.
3. Vibhute A, Bodhe SK. Applications of image processing in agriculture: a survey. International Journal of Computer Applications. 2012 Jan 1; 52(2):34-40.
4. Pratiwi D. The use of self organizing map method and feature selection in image database classification system. 2012 Jun 1; p.1-5.
5. Xin D, Bingfang W, Jihua M, Qiangzi L, FeiFei Z. A Method to Assess Land Productivity in Huang-Huai-Hai Region Using Remote Sensing. IEEE, 2010 International Conference on Multimedia Technology (ICMT). 2010 Oct 29; p. 1-5.
6. Ballabio D, Vasighi M. A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies. Chemometrics and Intelligent Laboratory Systems. 2012 Aug 15; 118:24-32.
7. Goswami SB, Saxena A, Bairagi GD. Remote Sensing and GIS based wheat crop acreage estimation of Indore district, MP. International Journal of Emerging Technology and Advanced Engineering. 2012; 2(3):200-3.
8. Defourny P, Blaes X, Bogaert P. Respective contribution of yield and area estimates to the error in crop production

forecasting. ISPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates 2007.

9. Chopin P, Blazy JM, Dore T. A new method to assess farming system evolution at the landscape scale. Agronomy for Sustainable Development. 2015 Jan 1; 35(1):325-37.
10. Grunwald S, McSweeney K, Rooney DJ, Lowery B. Soil layer models created with profile cone penetrometer data. Geoderma. 2001 Sep 30; 103(1):181-201.
11. Baskar SS, Arockiam L, Jeyasimman L. SOM-ANN A novel approach to analyzing the Brown Plant Hopper. International Journal of Computer Science and Applications (TIJCSA). 2013 Jun; 2(04):107-16.
12. Bingfang W, Qiangzi L. Crop area estimation using remote sensing on two-stage stratified sampling. International Society for Photogrametry and Remote Sensing (ISPRS). 2004 Jul; 20:12-23.
13. Xiao X, Boles S, Frolking S, Li C, Babu JY, Salas W, Moore B. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sensing of Environment. 2006 Jan 15; 100(1):95-113.
14. Cao X, Li Q, Du X, Zhang M, Zheng X. Exploring effect of segmentation scale on orient-based crop identification using HJ CCD data in Northeast China. IOP Publishing: IOP Conference Series: Earth and Environmental Science. 2014; 17(1):012047.
15. Hu X, Weng Q. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. Remote Sensing of Environment. 2009 Oct 31; 113(10):2089-102.
16. Shao Y, Fan X, Liu H, Xiao J, Ross S, Brisco B, Brown R, Staples G. Rice monitoring and production estimation using multitemporal RADARSAT. Remote sensing of Environment. 2001 Jun 30; 76(3):310-25.
17. Suma VR, Renjith S, Ashok S, Judy MV. Analytical Study of Selected Classification Algorithms for Clinical Dataset. Indian Journal of Science and Technology. 2016 Mar 22; 9(11):1-9.
18. Sudarshan MR, Jayapradha A, Amarnath DJ. Evaluation of Groundwater Quality at Oragadam -A GIS Approach. Indian Journal of Science and Technology. 2016 Apr 27; 9(14):1-8.