CROP YIELD PREDICTION SYSTEM USING MACHINELEARNING

1. MIRIYALA. BABYVIDYASRI.

PG Scholar, Department of Computer Science,

SVKP & Dr K S Raju Arts & Science College,

Penugonda, W.G.Dt., A.P, India

vidyamiriyala@gmail.com

2. K. LAKSHMANA REDDY

Associate Professor in Computer Science,

SVKP & Dr K S Raju Arts & Science College,

Penugonda, W.G.Dt., A.P, India.

Klreddy29@yahoo.co.in

ABSTRACT

Agricultural monitoring, particularly in developing countries, can aid in the prevention of famine and the support of humanitarian efforts. One major challenge is yield estimation, which is the prediction of crop yields prior to harvesting. Using publicly available remote sensing data, we present a scalable, accurate, and low-cost method for predicting crop yields. This solution, if implemented at the government established soil health centers, could assist all farmers in using minimal fertilizers in order to maintain soil health while also providing them with the opportunity to gain the most revenue from the same piece of land.

As a result, it would be a win-win situation for all parties involved. This is accomplished through the use of technologies such as Machine Learning and Image Processing.

The Machine Learning algorithm is used for prediction analysis, which includes recommending the best crop as well as the corresponding bio-fertilizer. Image Processing provides a technological foundation that could be used for future development projects in the field of automated drones or tractors because it generates the shortest number of turns through the field.

INTRODUCTION

1.1 Introduction

According to FAO (2015), 795 million people still lack an adequate food supply, and by 2050, there will be two billion more people to feed (Dods and Bartram 2016). Ending hunger and improving food security are primary goals of the United Nations 2030 Agenda for Sustainable Development (United Nations 2015). Yield estimation, or the ability to predict crop yields well before harvesting, is a critical challenge in addressing food security issues. In the face of climate change and droughts, agricultural monitoring, particularly in developing countries, can improve food production and aid humanitarian efforts (Dods and Bartram 2016). Existing approaches to crop yield modelling rely on survey data and other crop growth variables (such as weather and soil properties).

This approach has proven to be extremely effective in the United States, where data is plentiful and of relatively high quality. Comprehensive surveys of weather parameters, such as Day met (Thornton et al. 2014), and land cover types, such as the Cropland Data Layer (Borman et al. 2011), are publicly available and greatly aid crop yield prediction. However, data on weather, soil properties, and precise land cover are typically unavailable in developing countries, which have the greatest need for accurate crop yield prediction. Remote sensing, on the other hand, is a globally accessible and cost-effective data source that has recently gained popularity. It is frequently used in computational sustainability applications such as species distribution modelling (Fink, Damola's, and Dave 2013; Kelling et al. 2012), poverty mapping (Xie et al. 2015; Edmon et al. 2015), climate modelling (Rostovsky et al. 2013), and natural disaster prevention (Rostovsky et al. 2013). (Boulton, Totton, and Williams 2016).

These multispectral remote sensing images, which include information other than the traditional visible wavelengths (RGB) and have a relatively high spatial and temporal resolution, contain a wealth of information on vegetation growth and thus agricultural outcomes. However, because the data is high-dimensional and unstructured, extracting useful features is difficult.

In this paper, we propose a method based on modern feature learning ideas that have recently resulted in massive improvements in a variety of computer vision tasks (Izhevsk, Subsieve, and Hinton 2012; Apathy et al. 2014). We overcome the lack of training data by using a novel dimensionality reduction technique. We specifically treat raw images as histograms of pixel counts and use a mean-field assumption to approximate the high dimensional histogram. On the basis of these histograms, deep learning architectures are trained to predict crop yield.

While this method works well, it does not explicitly account for the spatiotemporal dependencies between data points. We overcome this limitation by layering deep models with Gaussian Process. We assess our method for forecasting county-level soybean production in the United States.

Our model outperforms competing techniques by a wide margin, while remaining interpretable in terms of feature importance, according to experimental results.

2. LITERATURE SURVEY

Crop yield forecasting using remotely sensed vegetation indices and crop phenology metrics

We developed empirical models predicting maize and soybean yield in the Central United States using data from NASA's

Moderate Resolution Imaging Spectroradiometer (MODIS) in conjunction with county-level data from the United States Department of Agriculture (USDA). We also tested MODIS' ability to capture inter-annual variability in yields as part of our analysis. Our findings show that the MODIS two-band Enhanced Vegetation Index (EVI2) is a better predictor of maize yields than the widely used Normalized Difference Vegetation Index (NDVI). The addition of crop phenology information derived from MODIS significantly improved model performance within and across years. Surprisingly, using moderate spatial resolution data from the MODIS Land Cover Type product to identify agricultural areas did not degrade model results when compared to using USDA higher-spatial resolution crop-type maps. Correlations between vegetation indices and yield were strongest for maize 65–75 days after Greenup and for soybeans 80 days after Greenup. The EVI2 was the best index for predicting maize yield in non-semi-arid counties ($R2 = 0.67$), but the Normalized Difference Water Index (NDWI) performed better in semi-arid counties ($R2 = 0.69$), likely because the NDWI is sensitive to irrigation in semi-arid areas with low agricultural density. NDVI and EVI2 predicted soybean yield equally well ($R2 =$ 0.69 and 0.70, respectively). Furthermore, in 2005, EVI2 was the best at capturing large negative anomalies in maize yield ($R2 = 0.73$). Overall, our findings indicate that using crop phenology and a combination of EVI2 and NDWI can significantly improve remote sensing-based maize and soybean yield models.

The global state of food insecurity. Meeting the international hunger targets for 2015.

Globally, 795 million people are undernourished, a decrease of 167 million over the last decade and 216 million fewer than in 1990–92. Despite significant population growth, the decline is more pronounced in developing regions. Slower and less inclusive economic growth, as well as political instability in some developing regions, such as Central Africa and Western Asia, have hampered progress in recent years.

The monitoring period for the Millennium Development Goal targets ends in 2015. The proportion of undernourished people in the developing world as a whole has fallen from 23.3 percent in 1990–92 to 12.9% today.

Some regions, such as Latin America, Asia's east and southeast, the Caucasus and Central Asia, and Africa's northern and western regions, have made rapid progress. Progress was also made in southern Asia, Oceania, the Caribbean, and southern and eastern Africa, but it was too slow to meet the MDG 1c target of halving the proportion of chronically undernourished people.

72 developing countries out of 129, or more than half of those monitored, have met the MDG 1c hunger target. Most experienced stable political and economic conditions, which were frequently accompanied by social protection policies aimed at vulnerable population groups.

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

In the remote sensing community, remote sensing data has been widely used to predict crop yield (Bolton and Friedl 2013; Johnson 2014). However, all existing approaches that we are aware of rely on hand-crafted features, assuming that they can capture the majority of the information about vegetation growth contained in high dimensional images. Normalized Difference Vegetation Index (NDVI) (Quarmby et al. 1993; Johnson 2014), two band Enhanced Vegetation Index (EVI2) (Bolton and Friel 2013), and Normalized Difference Water Index (NDWI) are some commonly used features (Satir and Barbarous 2016).

Disadvantages:

- While much effort has been put into feature engineering, existing features are fairly crude indexes that rely on a small number (usually two) of the available image bands.
- Second, in existing approaches, high-order moments of the features are rarely explored. In most cases, the regression output is ground truth average yield data over a region, with features provided as input for all locations within that region. Most studies either compute the mean (first moment) of the features over the region of interest (Johnson 2014) or conduct sampling (Kuwata and Shibasaki 2015).

3.2 Proposed System:

In contrast to previous approaches, we are the first to use modern representation learning ideas from AI to automatically discover relevant features from raw data, inspired by recent successes in computer vision and speech recognition. Our experimental results indicate that our learned features are far more effective, and that bands that are typically overlooked may play an important role.

Advantages:

- Farmers can determine which crops are feasible based on their soil type;
- Farmers have a chance of increasing their income based on analysis.

3.3 System Modules Dataset collection:

At this stage, a data set with temperature, humidity, and potassium is created. Nitrogen, phosphorus, labels 1–7, and crop information

Preprocessing Algorithm training:

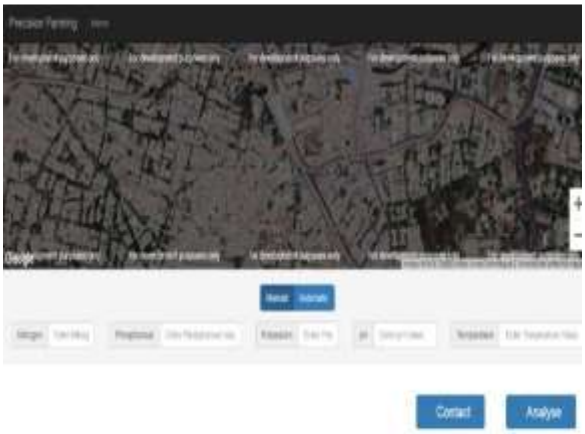
At this stage, data is collected from the dataset and divided into testing and training groups before being fed into the algorithm and fitted to it.

User Module:

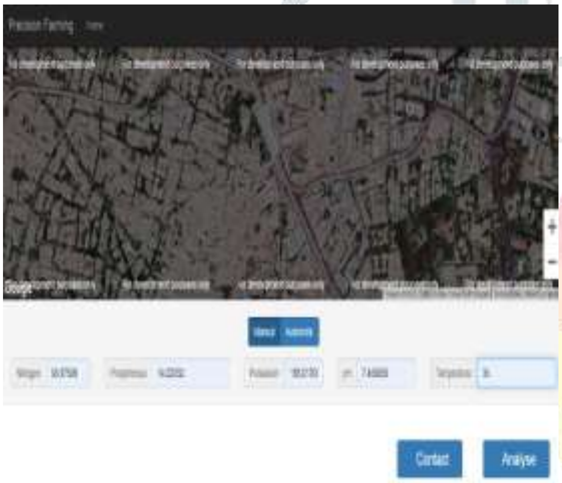
At this stage, the user enters all of the website's features and receives an output indicating which crop is best and the yield for each crop.

4. RESULTS

Main page

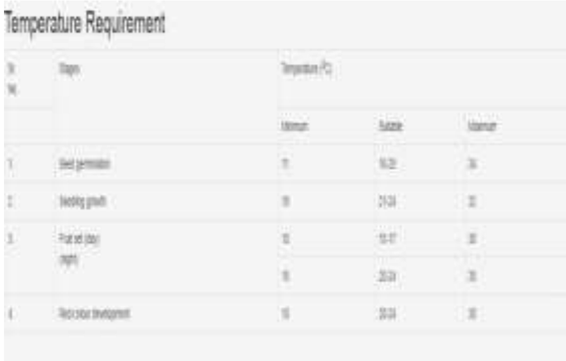


Enter Required values for manual analysis:



View Predicted values:

Temperature Requirement:



5. CONCLUSION

Based on low-cost remote sensing data, this paper presents a deep learning framework for crop yield prediction. It enables Realtime forecasting throughout the year and is applicable globally, particularly in developing countries where field surveys are difficult to conduct. We are the first to use modern representation learning ideas for crop yield prediction, and we successfully learn much more effective features from raw data than the traditional hand-crafted features. We present a Deep Gaussian Process framework that successfully removes spatially correlated error and propose a dimensionality reduction approach based on histograms, which may inspire other applications in remote sensing and computational sustainability. The model provides cutting-edge prediction accuracy and will have a significant impact on sustainable agriculture and food security.

Future Enhancements:

It is not possible to create a system that meets all of the user's requirements. As the system is used, the user requirements change. The following are some future enhancements that can be made to this system:

- As technology advances, it is possible to upgrade the system and adapt it to the desired environment.
- Security can be improved using emerging technologies such as single sign-on based on future security issues.

REFERENCES

1. Bolton, D. K., and Friedl, M. A. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology 173:74–84.
2. Boryan, C.; Yang, Z.; Mueller, R.; and Craig, M. 2011. Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program.
3. Geocarto International 26(5):341–358. Boulton, C. A.; Shotton, H.; and Williams, H. T. 2016. Using social media to detect and locate wildfires. In Tenth International AAAI Conference on Web and Social Media.
4. Breiman, L. 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author).

Statist. Sci. 16(3):199–231.

5. Chu, W., and Garamendi, Z. 2005. Preference learning with gaussian processes. In Proceedings of the 22nd international conference on Machine learning, 137–144. ACM.
6. Cressy, N., and Hawkins, D. M. 1980. Robust estimation of the variogram: I. Journal of the International Association for Mathematical Geology 12(2):115–125.
7. DAAC, N. L. 2015. The modis land products. http://lpdaac.usgs.gov. Dodds, F., and Bartram, J. 2016. The Water, Food, Energy and Climate Nexus: Challenges and an Agenda for Action. Routledge.

8. Ermon, S.; Xue, Y.; Toth, R.; Dilkina, B.; Bernstein, R.; Damoulas, T.; Clark, P.; DeGloria, S.; Mude, A.; Barrett, C.; and Gomes, C. 2015. Learning largescale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in East Africa. In AAAI Conference on Artificial Intelligence.

9. FAO. 2015. The state of food insecurity in the world. Meeting the 2015 international hunger targets: Taking stock of uneven progress.

10. Fink, D.; Damoulas, T.; and Dave, J. 2013. Adaptive spatiotemporal exploratory models: Hemispherewide species distributions from massively crowdsourced ebird data. In AAAI.

ABOUT AUTHORS:

MIRIYALA. BABYVIDYASRI is currently pursuing MCA in SVKP & Dr K S Raju Arts & Science College, affiliated to Adikavi Nannaya University,

Rajamahendravaram. Her research interests include Data Structures Web Technologies, Operating Systems and Artificial Intelligent.

K. Lakshmana Reddy is working as an Associate professor in the Department of Computer Science in SVKP & Dr K S Raju Arts &Science College, Penugonda, A.P. He received MCA from Andhra University, ‘C’ level from DOEACC, New Delhi and M.Tech from Acharya Nagarjuna University, A.P. He attended and presented papers in conferences and seminars. He has done online certifications in several courses from NPTEL. His areas of interests includes Computer Networks ,Network Security and Cryptography, Formal Languages and Automata Theory and Object Oriented programming languages.