Machine Learning in Soil Classification and Crop Detection

Rafega Beham
New Horizon College of Engineering

Machine Learning in Soil Classification and Crop Detection

Ashwini Rao[1] Janhavi U[2] Abhishek Gowda N S[3] Manjunatha[4] Mrs.Rafega Beham A[5]
[1,2,3,4,5]Department of Information Science and Engineering
[1,2,3,4,5]New Horizon College of Engineering, Bangalore, India

Abstract— This paper describes the SVM based classification and grading of soil samples using different scientific features. Different algorithms and filters are developed to acquire and process the colored images of the soil samples. These developed algorithms are used to extract different features like color, texture, etc. Different soil types like red, black, clay, alluvial, etc are considered. The classification makes use of Support Vector Machine, machine learning technique. SVM seeks to fit an optimal hyper plane between the classes and uses only some of the training samples that lie at the edge of the class distributions in feature space (support vectors). This should allow the definition of the most informative training samples prior to the analysis. The accuracy of a supervised classification is dependent to a large extent on the training data used. Till now classification of soil and classification of crop for the appropriate soil is done separately. This project aims at combining both the techniques, where classification of crop for appropriate soil is a part of classification of soil.

Key words: Crop Detection, Soil Classification

I. INTRODUCTION

In past years, automation and intelligent sensing technologies have revolutionized each and every part of our world like agriculture, building, aeronautics etc. These initiatives have been accredited to the rising concerns about product quality and safety. Also, rising labour costs, shortage of skilled workers and the need to improve production processes have all put pressure on producers and processors. Automated solutions are the answer for the problems that are being faced today by the agriculture world. Agriculture planning plays a significant role in economic growth and food security of agro-based country. Selection of crop(s) is an important issue for agriculture planning. Crop selector could be applicable for minimize losses when unfavourable conditions may occur and this selector could be used to maximize crop yield rate when potential exists for favourable growing conditions. Machine Vision Systems (MVS) provide an alternative to manual inspection of soil samples for their characteristic properties and the amount of nutrient material. Computers have been successfully employed for classification of plants, diagnosis of plant diseases, recognition of leaves and soil, etc. Till now classification of soil and crops were done separately. But in this methodology two techniques are implemented together for providing a great usage for farmers and agriculture industry.

II. RELATED WORK

In [1] author describes the measurement of soil moisture is an important issue in the research and application of agricultural wireless sensor networks. The research on agricultural wireless sensor network has practical and profound significance on the popularization of agricultural Internet of things and the realization of modern agriculture. However, the existing agricultural wireless sensor network still has the problem of high cost on agricultural sensors, which seriously restricts the development of agricultural wireless sensor network. Take soil moisture sensors for example, the existing methods to measuring soil moisture contain direct methods and indirect methods. Direct methods refer to those methods to measuring the soil moisture by drying soil sample in order to estimate the exact soil moisture, which are standard methods with great effort. Indirect methods refers to those methods to using other physical quantities related to soil moisture including the neuron probe, gamma ray, resistance, time-domain reflect meter(TDR), microwave remote sensing and so forth so as to estimate the actual soil moisture. Among them, the neuron probe and gamma ray have high cost and danger. High cost and complexity of soil moisture sensor become the bottleneck of promotion of agricultural wireless sensor network. In this way, a soil moisture classification model based on Support Vector Machine (SVM) with low-cost sensors as a way of substitution for high-cost soil moisture sensors is proposed by the author to meet the practical needs of agricultural wireless sensor network.

In [2] author describes the prediction of yield rate of crop, weather prediction, soil classification, crop classification and crop disease detection for agriculture planning using statistics methods or machine learning techniques. This research proposed a technique named Crop Selection Method (CSM) to solve crop selection problems and to maximize the yield rate of crop over season and to achieve maximum economic growth of the country in agriculture sector. Crop selector could be applicable for minimize losses and to maximize crop yield rate when favourable conditions apply. There are two factors that influence the crop yield rate: first is seed quality which can be improved using hybridization technology, second is crop selection method based on favourable conditions. Crop production rate depends on geography of a region, weather conditions, type of soil, soil composition and harvesting methods. Based on different parameters different prediction models were developed. These prediction models are classified into two different types: first is a traditional statistics model where it generates a global model over entire sample space. Second is machine learning technique which is emerging technology for knowledge mining that relates input and output variables which is hard to obtain statistically. IN traditional statistics methods, structures of data model needs to be assumed priory, where as machine learning techniques need not assume this structure. This is useful characteristics for machine learning techniques to model complex non-linear behaviour in crop yield prediction. Machine learning methods which are widely used in prediction technique are regression tree, support vector machine, k-nearest neighbours and artificial neural network. In this research, some machine learning techniques are studied and comparative analysis is presented.

In [3] author describes an approach for automating classification procedure. Firstly, Segmentation algorithm is applied to segment the measured signals. Secondly, the salient features of these segments are extracted using boundary energy method. Based on the measured data and extracted features classifiers to assign classes with the segments are built. This paper makes use of Decision trees, ANN and SVM. Cone Penetration Testing (CPT) methodology is used for classifying sub-surface soil. Feature extraction is performed using Boundary Energy Data by parameterising experts' perception of the shape of the series data. The overall classification scheme is based on CPT data. Soil Classification is done based on different algorithms for different soil types. In this paper a method to classify series data where the constraint of contiguity has to be maintained is presented.

III. PROPOSED SYSTEM

A. *Image Acquisition*

Different images of soil samples which are to be classified are captured using color camera and are provided as an input to the system. The features of each type of soil are collected and are stored in a separate database. This database is later used in the final stage for soil and crop detection.

Fig. 1: Images of Soil Samples

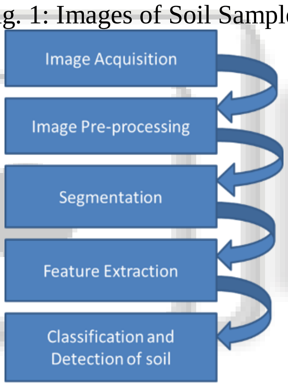

Fig. 2: Flowchart of the proposed system

B. *Image Pre-processing*

The image acquired from our previous stage is not error free. The quality of the image is decisive for results of analysis as it affects both the ability to detect features under analysis and precision of subsequent measurement. In order to get an error free image pre-processing techniques are applied. This phase is also known as the enhancement of the image since the image is enhanced by improving its contrast and removing errors to obtain a better quality image for our future processes. The image contains errors like noise or artifacts like scratches, lapping tracks, comet tails, etc which needs to be eliminated before the further processes. Hence a filter called the Smoothing filter is made use to remove the noise and artifacts from the image. There are two types of filters: low pass filter and high pass filters. Smoothing filter is a low pass filter. It is used to remove high spatial frequency noise from a digital image. Smoothing filter employs a moving window operator which affects one pixel of the image at a time, changing its value by some function of a local region of pixels. The operator moves over the image to affect all the pixels in the image. Thus with many iterations the smoothing filter gradually enhances the image by removing the errors. For more accuracy and clarity in the image its edges and contrast needs to be enhanced. Colour map of the image are used for enhancing edges and contrast.

C. *Segmentation*

Once the enhancement of the image is completed in the previous stage using image pre-processing techniques segmentation of the image is performed. Popular known algorithm K-means Clustering algorithm is employed for segmentation of the image. It is used as a partition clustering which aims at partitioning a given data set into disjoint subsets so that specific clustering criteria are optimized. The most widely used criterion is the clustering error criterion which for every point computes its squared distance from the corresponding cluster centre and then takes the sum of these distances for all points in the data set. The centre of the cluster is picked and each pixel of the image is assigned to the cluster. Re-computation of cluster centre is done by using the average of all the pixels. This iteration continues and the next step is to take each point belonging to a given data set and associate it to the nearest centroid.

K-means Algorithm is as follows-

1) Input: K, set of points $x_1........x_n$.
2) Place the centroids $C_1........C_k$ at random locations.
3) Repeat until convergence:

For each point x_i:
Find nearest centroid C_j arg min $D(x_i, C_j)$
Assign the point x_i to cluster j
For each cluster j=1.....K:

$$Cj(a) = \frac{1}{nj} xi -> Cj \sum xi(a) \quad \text{for each } a = 1 \ldots \ldots d$$

For each cluster j=1......K:
New centroid C_j= mean of all points x_i assign to cluster j in previous step.

4) Stop when none of the cluster assignments change.

D. *Feature Extraction*

After the segmentation of the image is performed by k-means in the above phase our next step is the feature extraction stage. This is the foremost step in this methodology. All the features that are required for us to classify the soil type and crop detection are done in this phase. A number of features like the texture, colour, intensity, saturation, hue, etc are extracted for detection of soil type. a filter known as Gabor Filter is implemented for feature extraction. Gabor Filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filter are similar to those of human visual system and they have been found to be particularly appropriate for texture representation and discrimination. A set of frequencies and orientation representations may be helpful for extracting useful features from the image. Also other features like entropy, standard deviance, mean, etc can be extracted using Gabor filter. The main and important feature of soil that is colour is needed to be extracted. Hence a measure called colour moments are used to differentiate images based on their features of colour. These provide a colour similarity between images which can be compared to the values of images indexed in the data base for tasks like image retrieval.

E. *SVM Classification*

Support Vector Machine (SVM) algorithm is used for soil classification. It has successful applications in many fields like bioinformatics, text, image recognition, etc. SVM is a universally accepted algorithm due to its simple nature. It is considered as an alternative to neural networks algorithm. The working nature is explained as follows: Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVM is effective to analyzing the separating planes and to identify the largest margin so that the support to the data points will be identified. So we want to learn the mapping: X->Y, where ‘x’ belongs to X is some object and ‘y’ belongs to Y is a class label.

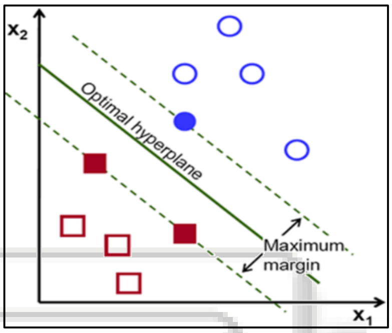

Fig. 3: classification by SVM

This algorithmic approach is based on distinctive characteristics analysis by analyzing the expected error minimization. This approach considered the empirical risk to improve the training procedure. The risk estimation is here based on the structural analysis so that the generalization error will be reduced. The error margin is analyzed under class deviation and based on it nearest training patterns are obtained. . This model is also based on the polynomial kernel representation so that the effective learning to the elements will be done and more accuracy will be obtained.

IV. Conclusion

The results from this study can be used for rapid identification of soil types when they arrive in railcars at the terminal soil elevators. Since this classification technique does not require time consuming image processing routines such as Fourier descriptors, it can readily be implemented using commercial imaging libraries with Digital Signal Processing (DSP) boards for real time operations. The work carried out has relevance to real world classification of soils and it involves both image processing and pattern recognition techniques. SVM involves the sound theory first and then implementation and experiments. A significant advantage of using SVM that the solution to an SVM is global and unique, they have simple geometric interpretations and gives sparse solution. The computational complexity of SVMs does not depend on the dimensionality of the input space and they are less prone to overfitting.

References

[1] Xiang Gao, Tancheng Lu, Peng Liu, Qiyong Lu,(2014), “A Soil Moisture Classification Model Based on SVM Used in Agricultural WSN”, Conf: on Information Technology and artificial Intelligence, Shanghai, China, pp.432-436.

[2] Kumar. R, Singh.M.P, Kumar.P, Singh.J.P, “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, (2015), Conf: on ICSTM, pp.138-145.

[3] B. Bhattacharya, and D.P. Solomatine, "An algorithm for clustering and classification of series data with constraint of contiguity", Proc. 3T"d nt. Conf: on Hybrid and Intelligent Systems, Melboume, Australia, 2003, pp. 489-498.

[4] Soumya sree, Asst. Prof. A Rafega Beham, (2015), “Automated Diagnosis of Retina Images for Diabetic Patients based on BP and SVM”, proceedings of the International Journal of Computer Science and Mobile Computing, NHCE, Bangalore, pp.299-306.