Computers and Electronics in Agriculture 135 (2017) 269–279

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Classification of agricultural soil parameters in India

M.S. Sirsat [a], E. Cernadas [a], M. Fernández-Delgado [a,*], R. Khan [b]

[a] Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), Universidade de Santiago de Compostela, Rúa de Jenaro de la Fuente Domínguez, Campus Vida 15782, Santiago de Compostela, Spain

[b] Department of Commerce & Management Science, Maulana Azad College, 431005 Aurangabad (M.S.), India

ARTICLE INFO

Article history:
Received 19 July 2016
Received in revised form 22 December 2016
Accepted 23 January 2017
Available online 24 February 2017

Keywords:
Soil type
Machine learning
Random forest
Soil fertility
Fertilizer recommendation

ABSTRACT

One of the backbones of the Indian economy is agriculture, which is conditioned by the poor soil fertility. In this study we use chemical soil measurements to classify many relevant soil parameters: village-wise fertility indices of organic carbon (OC), phosphorus pentoxide (P_2O_5), manganese (Mn) and iron (Fe); soil pH and type; soil nutrients nitrous oxide (N_2O), P_2O_5 and potassium oxide (K_2O), in order to recommend suitable amounts of fertilizers; and preferable crop. To classify these soil parameters allows to save time of specialized technicians developing expensive chemical analysis. These ten classification problems are solved using a collection of twenty very diverse classifiers, selected by their high performances, of families bagging, boosting, decision trees, nearest neighbors, neural networks, random forests (RF), rule based and support vector machines (SVM). The RF achieves the best performance for six of ten problems, overcoming 90% of the maximum performance in all the cases, followed by adaboost, SVM and Gaussian extreme learning machine. Although for some problems (pH,N_2O,P_2O_5 and K_2O) the performance is moderate, some classifiers (e.g. for fertility indices of P_2O_5, Mn and Fe) trained in one region revealed valid for other Indian regions.

1. Introduction

According to data of year 2011, India devotes 60.5% of its land[1] to agriculture (CIA, 2016), distributed among arable land (52.8%), land for permanent crops (4.2%) and pastures (3.5%). Share of agriculture and related activities was 11.3% of the Gross State Domestic product (GSDP) in 2013–2014. Data from the Directorate of Economics and Statistics (2015) show that in year 2013–2014 the cultivation areas of major crops were 15 and 57 millions of hector in Kharif and Rabi seasons, respectively. However, agriculture in India is conditioned by the poor fertility of the soil, which depends on the levels of its nutrients. The physical, chemical and biological properties of the soil are useful to evaluate its fertility, to design a cultivation plan and to predict the crop productivity. The information technologies, and specifically machine learning (ML), offer new possibilities in the field of agriculture and may help in data evaluation for decision making. The geographical study area of the current paper is the region of Marathwada, in the state of Maharashtra, one of the most prominent agricultural states in India, located at 19° 52′ 59.88″ North and 75° 19′ 59.88″ East. Its soil is made of basalt rock with scarlet, blackish and yellowish colors. The soil classification is useful to maintain and enhance its productivity, to avoid soil degradation problems and to overcome environmental damage. The major challenge is to increase crop yield for solving global food security problem. However, soil quality and crop yield are negatively affected by changing trends of temperature and rainfall, insufficient water and light, agriculture practices and absence of nutrients. It is important to develop an effective nutrient management by means of an adequate soil analysis and a proper application of fertilizers. Hence the relevance of a research effort to classify soil parameters such as the fertility indices for several nutrients (OC,P_2O_5, Mn and Fe, among others), soil pH, soil type, preferred crop and levels of several nutrients (N_2O,P_2O_5 and K_2O) which are relevant for fertilizer recommendation. The interest of predicting the levels of these magnitudes with machine learning techniques is to avoid the need to chemically measure these magnitudes, thus reducing the cost of the analysis and saving time of specialized technicians. The current study tries to enhance the accuracy of soil problem interpretation for Indian agriculture, although similar studies would benefit other nations around the globe.

2. Related work

Several studies (Mucherino et al., 2009) have applied ML techniques to solve soil problems in agriculture, namely to predict soil fertility, defined as the soil ability to supply the required nutrient

* Corresponding author.
E-mail address: manuel.fernandez.delgado@usc.es (M. Fernández-Delgado).

[1] https://www.cia.gov/library/publications/the-world-factbook/fields/2097.html.

http://dx.doi.org/10.1016/j.compag.2017.01.019
0168-1699/

levels and water for high quality crop yield. The soil fertility was predicted using artificial neural networks (ANNs) with Levenberg-Marquadt based back-propagation (Sheela and Sivaranjani, 2015), and also using partial least squares regression (Obade and Lal, 2016) using as input data the soil bulk density, electrical conductivity (EC), available water capacity, soil OC, pewamo silty clay loam, glynwood silt loam, kibbie fine sandy loam, crosby silt loam and crosby celina silt loams soil. The crop yield has been predicted using clustering techniques (Narkhede and Adhiya, 2014). One-R, J48, K-nearest neighbors (KNN) and A priori classifiers have been used to predict wheat yield (Romero et al., 2013) using as inputs phenotypic plant traits (thousand grain weight, plant height, peduncle length, harvest index, spikelets number, grain number, grain weight and spike fertility). The wheat yield has also been quantified (Pantazi et al., 2016) as low, medium and high with supervised Kohonen and counter-propagation neural networks, and with XY-fusion models using multi-layer soil data: pH, moisture content, total nitrogen, total carbon, magnesium, calcium, cation exchange capacity, available phosphorus and satellite imagery crop growth characteristics. Decisions about insecticide application (either spray or non-spray) for leafroller pest monitoring on kiwifruit are recommended (Hill et al., 2014) using decision tree (DT), naive Bayes classifier, RF, adaboost, SVM and logistic regression (LR). The generalized regression neural network was used to forecast plant diseases (Chtioui et al., 1999) for leaf wetness prediction. The RF provides the best accuracy for mapping the soil class spatial distribution in three semi-arid study areas with different sets of environmental covariates (Brungard et al., 2015), compared to clustering algorithms, discriminant analysis, multinomial logistic regression (MLR), ANN, DT and SVM. The soil has also been classified in 11 orders and 18 great groups from satellite images at 100 m spatial resolution, using classification and regression trees (CART), bagging with CART base classifiers, RF, KNN, nearest shrunken centroid, ANN, MLR, logistic model trees and SVM (Heung et al., 2016).

Our work deals with several problems analyzed in some of the previous papers, such as prediction of soil type (although with a different set of classes) and fertility indices (which in our paper are specific for OC,P_2O_5, Fe and Mn). However, our study analyses more soil problems than the previous works, including nutrients N_2O,P_2O_5 and K_2O, which allows to develop a fertilizer recommendation, the soil pH and prediction of suitable crops. We also evaluate the validity of the classifiers trained in one region to model data from different regions. However, available data for the current paper do not allow to predict crop yield, insecticide application nor plant diseases, as in some previous papers. On the other hand, our data are exclusively chemical measurements (see Section 3), excluding data such as satellite images and phenotypic plant traits. Finally, we use a wider and more diverse collection of classifiers than previous papers, which are specifically selected due to the good performance that they exhibited in the experimental comparison (Fernández-Delgado et al., 2014).

3. Materials and methods

In the current research we use data collected from the region of Marathwada by the State Government of Maharashtra (India) during years 2011 to 2015. Details about calibration of each input magnitude are publically available.[2] The inputs that we use are the following soil parameters: N_2O, measured by the soil testing laboratories using alkaline permanganate (Jones, 1912; Subbaiah and Asija, 1956); OC, using carbon spectrophotometric (Bowman et al., 1991); pH, using pH meter method (Richards et al., 1954); EC, using EC meter method (Richards et al., 1954); K_2O, using flame photometric (Ford, 1954; Jackson, 1958); and P_2O_5, using the Olsen's method (Olsen, 1954). Micro nutrients as Fe, copper (Cu), zinc (Zn), Mn and boron (B), which are useful to evaluate imbalances in soil nutrients, are measured using atomic absorption spectroscopy (Leggeit and Argyle, 1983). The pH is expressed as the decimal logarithm of the Hydrogen concentration. The values of N_2O,K_2O and P_2O_5 are expressed in kilograms per hector (kg/ha), while Fe, Cu, Zn, Mn, B and SO_4 are expressed as parts per million (PPM). The EC is expressed in mili-Siemens per centimeter (mS/cm), while OC and $CaCO_3$ are expressed as mass percentages (denoted as %).

3.1. Classification problems

The following subsections describe the ten soil parameters which are classified in the current work.

3.1.1. Soil OC,P_2O_5, Mn and Fe village-wise fertility indices

There are proofs of the interconnection among organic matter, ecosystem sustainability and soil fertility (Feller et al., 2012), which is important for crop yield. This fertility mainly depends on OC,N_2O,P_2O_5 and K_2O, considered soil major nutrients because they appear in large quantities, and also on micro nutrients Fe, Mn, Zn and Cu, which appear in smaller quantities. However, our work is restricted to fertility levels of OC,P_2O_5, Fe and Mn, due to data availability. The OC is very important for the soil health, biological activity and crop productivity (Reeves, 1997), and adequate fertilizers help to keep its level (Turmel et al., 2015). The P_2O_5 is used by plants for cell signaling, phosphorylation and bioenergetics, while Fe and Mn help chlorophyll to absorb light energy for photosynthesis. The agriculture planning of the Indian Government requires to determine the village-wise fertility indices N_I for the previous nutrients, using the thresholds listed in Table 1 to quantify their levels as low, medium and high. For each village and nutrient, N_l, N_m and N_h are the number of patterns (i.e., cultivation lands) with low, medium and high levels. The village-wise fertility index N_I for a nutrient is calculated as $N_I = (N_l + 2N_m + 3N_h)/N_t$, being N_t is the total number of patterns analyzed for a village. The value of N_I (which is the same for all the patterns in the village) is then quantified into low, medium and high levels, according to the threshold values listed in the rightmost column of Table 1. The classification of the village-wise fertility index uses the inputs listed in the first four lines of Table 2. The labels OC-F, P_2O_5-F, Mn-F and Fe-F mean village-wise fertility index of OC,P_2O_5, Mn and Fe, respectively, whose values are completely different to the inputs OC,P_2O_5, Mn and Fe. Our data (see Table 2) only include patterns with OC-F and Fe-F (resp. Mn-F) in levels low and medium (resp. medium and high).

3.1.2. Classification of soil nutrients N_2O,P_2O_5 and K_2O

The direct measurement of soil N_2O is difficult, but it is largely present in the OC form (97–99%), so that it can be determined

Table 1
Intervals defined by the Department of Agriculture & Cooperation (2011) for the major and micro nutrients respectively (Muhr et al., 1965; Katyal and Rattan, 2003), and rate of nutrient index (Rammoorthy and Bajaj, 1969).

	Major nutrients		Micro nutrients		Index
	OC	P_2O_5	Mn	Fe	
	(%)	(kg/ha)	(PPM)		
Low <	0.5	10	1	2.5	1.67
Medium	0.5–0.75	10–24.6	1–2	2.5–4.5	1.67–2.33
High >	0.75	24.6	2	4.5	2.33

[2] http://mpkv.ac.in/Handler2.ashx?sel=Data&&tbl=DataMst&&whr=Type=%27PDF%27%20and%20Id=2004 (in Marathi language).