Classifying Soil Type Using Radar Satellite Images

Sajib Ahmed[1]
jack6148@gmail.com
Teresa Gonçalves[1]
tcg@uevora.pt
Luís Rato[1,2]
lmr@uevora.pt
Pedro Salgueiro[1]
pds@uevora.pt
J. R. Marques da Silva[3,4]
jmsilva@uevora.pt
Filipe Vieira[4]
fsvieira@agroinsider.com
Luis Paixão[4]
lpaixao.agroinsider@gmail.com

[1] Departmento de Informática,
Universidade de Évora, Portugal
[2] CIMA, Universidade de Évora, Portugal
[3] MED, Universidade de Évora, Portugal
[4] Agroinsider Lda., Évora, Portugal

Abstract

The growth of the crop is dependent on soil type, apart from atmospheric and geo-location characteristics. As of now, there is no direct and cost-free method to measure soil property or to classify soil type. In this work, we proposed a machine learning model to classify soil type using Sentinel-1 satellite radar images. Further, the developed classifier achieved 72.17% F1-score classifying sandy, free and clayish on a set of 65003 data points collected over one year (from Oct 2018 to Sep 2019) over 14 corn parcels near Ourique, Portugal.

Keywords: Remote Sensing, Soil Electrical Conductivity, Sentinel-1, Machine Learning, Random Forest

1 Introduction

Precision farming involves the collection of detailed information of mineral, nutrients, water, soil texture, cation exchange capacity, drainage conditions, organic matter level, salinity, and subsoil characteristics over farmland [3]. Over the last few decades, many new technologies have been developed for measuring soil properties, and one of such is using remote sensing techniques [2].

Sentinel-1 [7] is a synthetic aperture radar instrument (SAR) satellite that provides images in two different polarizations: VV (vertical transmit, vertical receive) and VH (vertical transmit, horizontal receive). It consists of a constellation of two satellites, Sentinel-1A and Sentinel-1B, which share the same orbital plane with a 12-day revisiting period.

In precision farming, detailed information about the spatial characteristics of farm operations like yield estimation, field attribute maps and forecasting harvesting date are made available to the farmer. This information is gathered using a wide array of electronic, mechanical and chemical sensors which leads to measure and map soil and plant properties. Soil Electro-Conductivity (EC) is one of the simplest, least expensive soil measurements available to precision farming today [8].

EC is the ability of a material to transmit (conduct) an electrical current and is usually expressed in miliSiemens/meter (mS/m). Soil EC is a measurement that characterizes soil properties which, in turn, affect the productivity of crops. These properties include water content, soil texture, soil organic matter (OM), depth to clay layer, the capacity of cation exchange (CEC), salinity, calcium and magnesium [4].

The objective of the present study is to build a classification model using machine learning algorithms that characterize soil types using Sentinel-1 radar images.

The rest of the paper is organized in the following sections: Section 2 introduces the data used in this work, while Section 3 describes the machine learning model, the experimental setup, experiments and results. Finally, Section 4 concludes the paper.

2 Data Set Construction and Characterization

The Electro-Conductivity value from a set of 14 parcels of corn fields (made available by Agroinsider [1]) was used as ground data points. These parcels are from Alentejo region with coordinates between (37°56'29.13" N , 8°22'21.95" W) and (37°55'32.44" N, 8°21'02.23" W). Figure 1 shows the Google View image of these 14 parcels. EC value was measured at 10-meter intervals resulting in a total of 65003 points.

Figure 1: Google view images of 14 parcels

Electro-conductivity real values were discretized, leading to three types of soil: sandy, free, and clayish. Table 1 presents the information about each type: the EC values interval and the number of points.

Soil Type	Value Range	Count
Sandy	$EC < 10mS/m$	24195
Free	$10mS/m \leq EC \leq 25mS/m$	31141
Clayish	$EC > 25mS/m$	9667

Table 1: Soil type information.

For each data point, along with the EC value, the respective latitude and longitude were also noted. With the collected coordinates, the corresponding values of VV and VH from the radar images were taken.

This radar data was collected from October 2018 to September 2019, the time span of one agricultural year. Since the Sentinel-1 revisiting time is 6 days, it resulted in a set of 60 pairs of values for each EC point measured. In this way, each soil point is characterized by 122 attributes: the soil type plus latitude, longitude and $60 * 2$ values of the radar images (60 dates and two polarizations: VV, VH). But latitude and longitude are not used as a parameter value in the ML algorithm.

Figure 2 represents the corresponding radar image for October 8, 2018 with VH polarization. And the variation of VH and VV value for one agriculture year (From Oct 2018 to Sep 2019) is shown in Figure 3.

3 Machine Learning Models

Three machine learning algorithms have been used to build classification models:

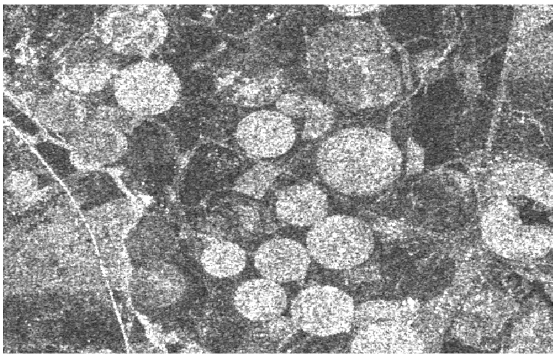

Figure 2: VH polarized radar image on 6^{th} October 2018

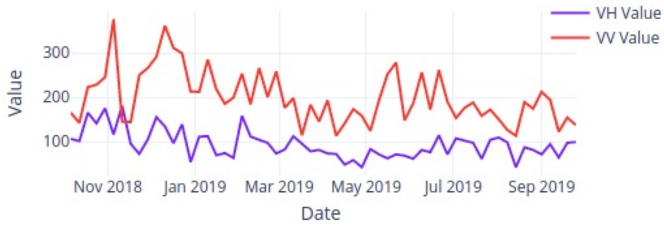

Figure 3: Variation of VH and VV values over a year in a specific point.

- Support Vector Machines (SVM) with a linear kernel.
- Random Forest (RF), a set of decision trees built from bootstrap samples of the training set where the candidate split in the learning process, is chosen from a random subset of the features.
- Extra Trees (ET), another ensemble classifier of decision trees, differs from RF in two points: each tree is trained using the whole learning sample and the top-down splitting in the tree learner is randomized.

3.1 Experimental Setup

A stratified train-test split was done over the dataset, with 80% for training (52002 samples) and 20% for testing (13001 samples).

We used Scikit-learn library [6] and RandomizedSearchCV [5] approach with 5-folds cross-validation to fine-tune the algorithms over micro-F1 measure. Parameters that produces the best results were: $nestimators = 189$, $max\_features = sqrt$, $max\_depth = 32$, $min\_samples\_split = 2$, $bootstrap = False$, $min\_samples\_leaf = 1$, and $criterion = gini$.

3.2 Experiments and Results

In order to evaluate the performance of the algorithms in this problem as well as the most relevant set of attributes, several experiments were carried out in a total of 153:

1. Algorithms: SVM, RF, ET
2. Time interval
   (a) 12 months
   (b) 3 months (Oct – Dec, Jan – Mar, Apr – Jun, Jul – Sep)
   (c) 1 month (Oct, Nov, Dec, Jan, Feb, Mar, Apr .... Sep)
3. Polarization: VV, VH, VV + VH

These preliminary results made it possible to draw the following conclusions:

- Data set of 12 months time interval shows better results in performance measures: precision, recall and F1-Score.
- Compared to the other shorter intervals, performance increase between 2% to 3% in the F1-score measure, when compared to the results obtained with the April-June interval. The April-June interval presents the 2nd best F1-score values.
- The performance measure using only one of the polarization is similar. But some are gain (between 2% and 7% in the F1-score measure) when using both polarizations.
- Random Forest present the outperform than others based on the performance measures.

Table 2 details the results using Random Forest for the time span of 12 months. It presenting the best results in the three performance measures. So from 12 months time interval, several conclusions can be drawn from

Soil Type	**Precision** (%)	**Recall** (%)	**F1-Score** (%)
Sandy	79.70	70.15	74.62
Free	68.25	84.76	75.62
Clayish	80.17	41.21	54.44

Table 2: Performance of the Random Forest model over the test set.

the results:

1. it is possible to observe that the model behaves reasonably for sandy and free soils; precision is about 10% higher for sandy soils (almost 80%) but, on the other hand, free soils present 15% higher recall (about 85%);
2. concerning clayish soils, a high precision (about 80%) is obtained at the expense of a significantly low recall (about 41%); this difference affects F1-score, which fails to reach 55%, while for other types of soil the value is around 75%;

4 Conclusions and Future Work

This work presents a machine learning model to classify soil type using Sentinel-1 satellite images. The developed model, using Random Forests, is able to achieve 74.62%, 75.62% and 54.44% F1-score for sandy, free and clayish soils, respectively.

In future, to improve the results of this work, we will enlarge the dataset with more parcels having different crops, including more features from radar like the angle of incidence and timing for example.

Funding

This work was supported by NIIAA (Núcleo de Investigação em Inteligência Artificial em Agricultura) project, Alentejo 2020 program (reference ALT20-03-0247-FEDER-036981).

References

[1] Agroinsider an agricultural consulting company. `https://www.agroinsider.com/`, -. Accessed: 18 09 2020.

[2] Yufeng Ge, J Alex Thomasson, and Ruixiu Sui. Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science*, 5(3):229–238, 2011.

[3] Robert Dwight Grisso, Marcus M Alley, David Lee Holshouser, and Wade Everett Thomason. Precision farming tools. soil electrical conductivity. -, 2005.

[4] Vr Ouhadi and Amir Reza Goudarzi. Factors impacting the electro conductivity variations of clayey soils. *Iranian Journal of Science and Technology Transaction B-Engineering*, 2007.

[5] David Paper and David Paper. Scikit-learn classifier tuning from complex training sets. *Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python*, pages 165–188, 2020.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[7] Paul Snoeij, Evert Attema, Malcolm Davidson, Berthyl Duesmann, Nicolas Floury, Guido Levrini, Björn Rommen, and Betlem Rosich. The sentinel-1 radar mission: Status and performance. In *2009 International Radar Conference" Surveillance for a Safer World"(RADAR 2009)*, pages 1–6. IEEE, 2009.

[8] Iulian-Florin Voicea, Mihai Matache, and Valentin Vladut. Researches regarding the electro-conductivity determination on different soil textures from romania, before sowing. *Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Agriculture*, 66(1), 2009.