AI in Agriculture

Miloš A. Kovačević[1], Petar M. Bursać[1], Branislav J. Bajat[1], Milan S. Kilibarda[1]

[1] Faculty of Civil Engineering, University of Belgrade, Serbia
Email: milos@grf.bg.ac.rs, pbursac@grf.bg.ac.rs, bajat@grf.bg.ac.rs, kili@grf.bg.ac.rs

Abstract

Soil organic carbon represents the main nutrient source for crop yields, which is of great importance to agricultural production. This research investigates the usage of transfer learning-based neural network model to predict SOC values from geochemical soil parameters. The results on datasets representing five Europian countries showed that the model was able to capture the valuable information contained in grassland soil samples when predicting the SOC values in cropland areas.

Key words: soil organic carbon (SOC), transfer learning, neural network

1 Introduction

Measuring and assessment of some soil components and properties is generally a time-consuming and costly procedure. The absence of measured data is often reimbursed by results of predictions or modeling [1]. A commonly used approach to the estimation of such soil parameters is based on their indirect assessment, using other already available soil parameters. One of the most important soil parameters is soil organic carbon (SOC). SOC is a vital part of the global carbon cycle and represents the main nutrient source for crop yields, which is of great importance to agricultural production [2]. Traditional SOC measurements are time-consuming and laborious. Therefore, using the AI approach, especially machine learning techniques, to estimate SOC from other geochemical parameters, is the direction of future development. SOC concentration depends on various factors such as soil type, climate, topography, and soil management practices. SOC is greatly influenced by vegetation through organic matter input and, consequently, land-use change is one of the most important factors which impacts SOC stock increase/decrease.

This paper examines the possibility of a neural network model to predict SOC in arable cropland from other geochemical parameters measured in different, but related land cover (grassland) areas, by means of instance-based transfer learning [3].

2 Instance-based non-inductive transfer learning

The proposed model for predicting SOC is designed to use the instance-based non-inductive transfer learning [3]. We first define the notions of the domain, task and transfer learning. A domain $\mathbb{D}$ consists of two compenents: a feature space $\mathscr{X}$ from which samples $\mathbf{x} \in \mathscr{X}$ come from, and a marginal probability distribution P_X that produces each sample $\mathbf{x}$. All samples are, or can be transformed to, vectors of real numbers $\mathbf{x} \in \mathbb{R}^n$. Given a domain of interest $\mathbb{D}(\mathscr{X}, P_X)$, a task $\mathbb{T}$ consists of two components: a label space $\mathscr{Y}$, and a predictive function $f : \mathscr{X} \to \mathscr{Y}$ for which $y = f(\mathbf{x})$. In a regression task,

$y \in \mathscr{Y}$ is a real number, while in a classification task it takes one of several discrete values (classification labels). It is common to interpret f as a probability $P(y|\mathbf{x})$, leading to $\mathbb{T} = \mathbb{T}(\mathscr{Y}, P_{Y|X})$. Often, there are two domains of interest, source domain $\mathbb{D}_s$ and target domain $\mathbb{D}_t$. They are represented with source-labeled and destination-labeled datasets where the labelling process is the outcome of two tasks $\mathbb{T}_s$ and $\mathbb{T}_t$: $(\mathbf{x}_{s_i}, y_{s_i})_{i=1\cdots n}$, where $\mathbf{x}_{s_i} \in \mathscr{X}_s$, $y_{s_i} \in \mathscr{Y}_s$, and $(\mathbf{x}_{t_i}, y_{t_i})_{i=1\cdots m}$, where $\mathbf{x}_{t_i} \in \mathscr{X}_t$, $y_{t_i} \in \mathscr{Y}_t$. According to [4], transfer learning can be defined as:

Definition 1: Given $\mathbb{D}_s, \mathbb{T}_s, \mathbb{D}_t$, and $\mathbb{T}_t$, transfer learning aims to help imporve the learning of the predictive function f_t from the target domain, using the knowledge in $\mathbb{D}_s$ and $\mathbb{T}_s$, where $\mathbb{D}_s \neq \mathbb{D}_t$ or $\mathbb{T}_s \neq \mathbb{T}_t$.

From Definition 1 follows that, for $\mathbb{D}_s = \mathbb{D}_t$ and $\mathbb{T}_s = \mathbb{T}_t$, a learning setting becomes a traditional machine learning problem. However, the nature of the difference between the domains or between the tasks can be used to categorize different transfer learning settings. We now define the instance-based non-inductive transfer learning setting:

Definition 2: Let $\mathbb{T}_s = \mathbb{T}_t$, that is $\mathscr{Y}_s = \mathscr{Y}_t$ and $P^s_{Y|X} = P^t_{Y|X}$. If $\mathscr{X}_s = \mathscr{X}_t$ and $P^s_X \neq P^t_X$ (i.e. $\mathbb{D}_s \neq \mathbb{D}_t$), the learning setting becomes instance-based and non-inductive.

Instance-based non-inductive setting assumes the same feature and label spaces, and the same underlying process that maps inputs to outputs in both domains. However, the marginal probability distributions of instances (samples) are different across domains. We assume the marginal probability distributions of the observed samples are different across various land cover types. Therefore, this setting can be applied when one tries to predict cropland OC values using the geochemical and OC values from grassland samples and only geochemical from cropland samples.

Suppose that $\mathbb{D}_s$, $\mathbb{T}_s$, and $\mathbb{D}_t$ are represented with $(\mathbf{x}_{s_i}, y_{s_i})_{i=1\cdots n}$, and $(\mathbf{x}_{t_i})_{i=1\cdots m}$. We would like to find the optimal parameters $\boldsymbol{\theta}^*_t$ of the target task prediction model under the assumption of the instance-based non-inductive setting. Using the empirical risk minimization framework [5], we minimize the following expectation:

$$\boldsymbol{\theta}^*_t = \arg\min_{\boldsymbol{\theta}_t} \mathbb{E}_{(\mathbf{x},y)\sim P^t_{X,Y}} \left[l(\mathbf{x}, y, \boldsymbol{\theta}_t) \right] \tag{1}$$

where $l(\mathbf{x}, y, \boldsymbol{\theta}_t)$ is a loss function defined for the target task. Using the definition of expectation and the Bayes' rule, (1) becomes:

$$\boldsymbol{\theta}^*_t = \arg\min_{\boldsymbol{\theta}_t} \mathbb{E}_{(\mathbf{x},y)\sim P^s_{X,Y}} \left[\frac{P_t(\mathbf{x}, y)}{P_s(\mathbf{x}, y)} l(\mathbf{x}, y, \boldsymbol{\theta}_t) \right] \tag{2}$$

Since, by Definition 2, $P^s_{Y|X} = P^t_{Y|X}$, and after using the Bayes' rule, (2) becomes:

$$\boldsymbol{\theta}^*_t = \arg\min_{\boldsymbol{\theta}_t} \mathbb{E}_{(\mathbf{x},y)\sim P^s_{X,Y}} \left[\frac{P_t(\mathbf{x})}{P_s(\mathbf{x})} l(\mathbf{x}, y, \boldsymbol{\theta}_t) \right] \tag{3}$$

Optimal parameters of the target model cannot be found by (3) since the expectation of the joint distribution in the source population is impossible to compute. The best we can do is to apply the empirical approximation on the training data:

$$\boldsymbol{\theta}^*_t = \arg\min_{\boldsymbol{\theta}_t} \sum_{i=1}^{n} \left[\frac{P_t(\mathbf{x}_{s_i})}{P_s(\mathbf{x}_{s_i})} l(\mathbf{x}_{s_i}, y_{s_i}, \boldsymbol{\theta}_t) \right] \tag{4}$$

Equation (4) suggests why this method is called "instance-based". Each source domain instance is weighted in the loss function with the ratio $\frac{P_t(\mathbf{x})}{P_s(\mathbf{x})}$, meaning that

if an instance is more probable to ocure in the target domain, then the optimization process pays more attention to it. If the probability ratio is 1 for all source instances, then the domains can be treated as one, and the loss function takes its standard form. The probability ratio can be estimated using the rejection sampling-based method for correcting sample selection bias [6]. This method introduces a new binary random variable $\delta \in \{0, 1\}$ which selects whether a sample will be accepted by the source domain or not: $P_s(\mathbf{x}) = P(\mathbf{x}|\delta = 1)$, and $P_t(\mathbf{x}) = P(\mathbf{x}|\delta = 0)$. The ratio can be calculated as follows:

$$\frac{P_t(\mathbf{x})}{P_s(\mathbf{x})} = \frac{P(\mathbf{x}|\delta = 0)}{P(\mathbf{x}|\delta = 1)} = \frac{P(\mathbf{x})P(\delta = 0|\mathbf{x})P(\delta = 1)}{P(\mathbf{x})P(\delta = 1|\mathbf{x})P(\delta = 0)} = \frac{P(\delta = 1)}{P(\delta = 0)}\left(\frac{1}{P(\delta = 1|\mathbf{x})} - 1\right) \quad (5)$$

Equation (5) suggests that the probability ratio is proportional to $\frac{1}{P(\delta = 1|\mathbf{x})}$. Hence, one can treat the evaluation of the ratio as a binary classification problem in which a classifier is trained to predict the probability of a sample being from the source domain ($P(\delta = 1|\mathbf{x})$), or the target domain ($P(\delta = 0|\mathbf{x})$).

3 Grassland-to-cropland SOC prediction model

In this research, $\mathbb{D}_s = \mathbb{D}_{grassland}$ and $\mathbb{D}_t = \mathbb{D}_{cropland}$. The feature spaces, $\mathscr{X}_{grassland} = \mathscr{X}_{cropland}$, of the proposed grassland-to-cropland SOC prediction model contain 5 physical and chemical properties measured at identical locations: *Nitrogen* (Total NCS in g/kg for < 2 mm soil fraction), *pH - H2O* (1:1 Soil-Water Suspension for < 2 mm soil fraction), *pH - CaCl2* (pH, CaCl2 Suspension for < 2 mm soil fraction), *Potassium* (Extractable in mg/kg for < 2 mm soil fraction), and *Electrical Conductivity* (Saturation Extract in dS/m for < 2 mm soil fraction). The label spaces, $\mathscr{Y}_{grassland} = \mathscr{Y}_{cropland}$, contain SOC values measured following the ISO 10694:1995 protocol [7]. Under the assumption of learning setting from Definition 2, domains and tasks are represented with $(\mathbf{x}_{grassland}, y_{grassland})_{i=1\cdots n}$, and $(\mathbf{x}_{cropland})_{i=1\cdots m}$.

The model is trained in two phases. In the first phase, a two-layer, fully-connected neural network classifier [8] is trained to distinguish between the grassland (source domain), and the cropland (target domain) samples. The activation function in each of the five hidden layer neurons is ReLU, while the output neuron performs the Sigmoid function. The network is trained to minimize the binary cross-entropy loss in a standard backpropagation procedure. When trained, the network assigns the probabilities of belonging to the grassland class to each sample from both land types. The assigned probabilities will be used in the next phase, to modify the mean squared error loss of the regression model according to (4) and (5).

The regression model uses a two-layer, fully-connected neural network with five hidden neurons and one linear output [8]. The network is trained in a standard backpropagation procedure. Optimal hyperparameters for both networks (learning rate, momentum, and the number of training epochs) were found in a standard cross-validation procedure.

4 Experimental model evaluation

The proposed model is evaluated using the subset of the LUCAS data set, which originally consists of 21 857 observations [9], with SOC ranging from 0.10 to 560.20 g/kg, as shown on the map in Figure 1. The subset included data from five countries: Sweeden (108, 153), Germany (410, 836), France (783, 1580), Austria (166, 117), and

Bulgaria (124, 255) − the numbers in parenthesis are the numbers of samples from the grassland and cropland land types respectively.

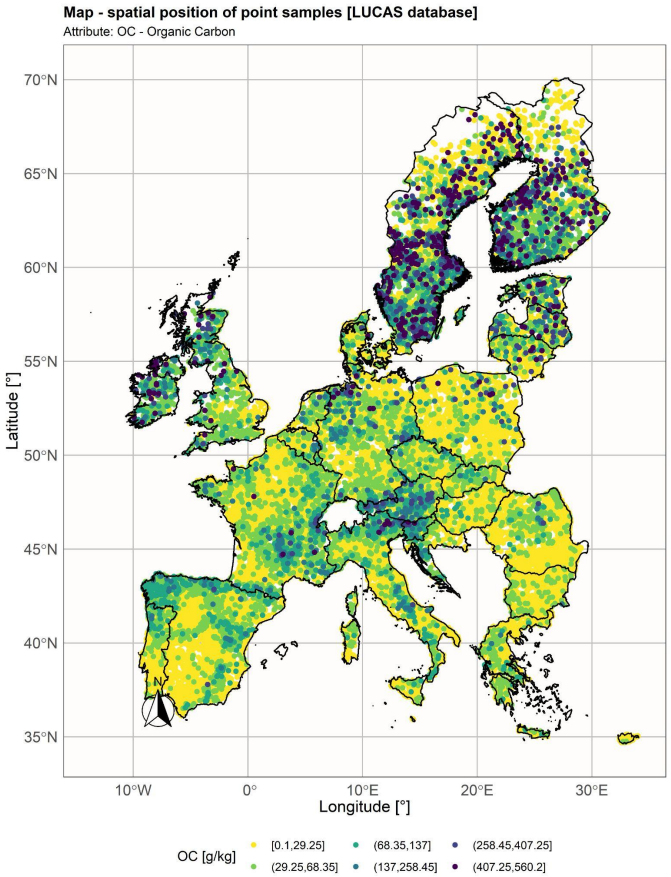

Figure 1: Map of spatial distribution of LUCAS point samples and SOC content values in [g/kg]

After training the transfer learning model on geochemical properties of grassland and cropland samples, and grassland SOC values, the prediction results on cropland SOC values were compared to classiscal setting in which we set $\frac{P_t(\mathbf{x})}{P_s(\mathbf{x})} = 1$. The proposed and the classical approach were compared using standard regression metrics, Normalized Root Mean Squared Error ($\frac{1}{\bar{y}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$), Normalized Mean Absolute Error ($\frac{1}{n\bar{y}}\sum_{i=1}^{n}|y_i - \hat{y}_i|$) and R^2 ($1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$). The results are shown in Table 1.

	Sweeden	Germany	France	Austria	Bulgaria
NRMSE Cl	.58	.39	.27	.20	.13
NRMSE TL	.54	.28	.26	.19	.14
NMAE Cl	.31	.26	.18	.15	.10
NMAE TL	.27	.17	.15	.14	.10
R2 Cl	.83	.78	.77	.93	.83
R2 TL	.85	.88	.80	.94	.83

Table 1: Comparing classical (Cl) and transfer learning (TL) approach: NRMSE and NMAE (lower the better), and R^2 (higher the better), indicate the benefits of the proposed approach.

All performance measures show that our transfer learning-based network behaves better than the classically trained network, with different levels of improvement among countries - the transfer of knowledge from grassland to cropland was quite successful in Germany and Sweeden, moderate in France, while in Austria and Bulgaria the differences were negligible. We suppose countries where the model achieved better results apply more appropriate agrotechnical measures to preserve SOC content in cropland (more cropland samples exhibit similar geochemical characteristics to typical grassland samples).

5 Conclusions

In this paper, an instance-based transfer learning model is created to predict SOC values of cropland samples using geochemical soil parameters and SOC values of grassland samples. Compared to a classical machine learning setting, using soil samples from five European countries, the model achieved better performance and showed the potential of transfer learning in regression problems in agriculture and soil science.

Acknowledgments

The research was funded by CERES project, by the Science Fund of the Republic of Serbia-Program for Development of Projects in the Field of Artificial Intelligence, with grant number 6527073.

References

[1] Kovačević M, Bajat B, and Gajić B, Soil type classification and estimation of soil properties using support vector machines. Geoderma, 154(3-4), 340-347, 2010.

[2] Obalum S. E, Chibuike G. U, Peth S, and Ouyang Y, Soil organic matter as sole indicator of soil degradation. Environmental monitoring and assessment, 189(4), 1-19, 2017.

[3] Yang Q, Zhang Y, Dai W, and Pan S, Transfer Learning. Cambridge: Cambridge University Press, 2020.

[4] Pan S, and Yang Q, A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345–1359, 2010.

[5] Vapnik V, Statistical Learning Theory. Wiley-Interscience, 1998.

[6] Zadrozny B, Learning and evaluating classifiers under sample selection bias. Proceedings of the Twenty-First International Conference on Machine Learning, 2004.

[7] Orgiazzi A, Ballabio C, Panagos P, Jones A, Fernández-Ugalde O, LUCAS Soil, the largest expandable soil dataset for Europe: A review. European Journal of Soil Science, 69(1): 140-153, 2018.

[8] Goodfellow I, Bengio Y and Courville A, Deep Learning. MIT Press, 2016.

[9] Jones A, Fernandez-Ugalde O, Scarpa S, LUCAS 2015 Topsoil Survey. EUR 30332 EN, Publications Office of the European Union: Luxembourg. 2020.