Computers and Electronics in Agriculture 144 (2018) 86–93

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China

Wei Wu[a,*], Ai-Di Li[b], Xin-Hua He[b,c], Ran Ma[b], Hong-Bin Liu[b], Jia-Ke Lv[a]

[a] College of Computer and Information Science, Southwest University, Beibei, Chongqing 400716, China
[b] College of Resources and Environment, Southwest University, Beibei, Chongqing 400716, China
[c] School of Biological Sciences, University of Western Australia, Crawley, WA 6009, Australia

ARTICLE INFO

Keywords:
AUC
Comparison
Kappa
Kernel function
Stratum

ABSTRACT

The variability of soil properties plays a critical role in soil and water conversation engineering. In this study, different machine learning techniques were applied to identify the soil texture classes based on a set of terrain parameters in a small mountainous watershed located in the core areas of Three Gorges of Yangtze River, southwest China. For this, the support vector machines (SVMs) with polynomial and Gaussian radius basis functions, artificial neural network, and classification tree methods were compared. The most commonly used performance measures including overall accuracy, kappa index, receiver operating characteristics (ROC), and area under the ROC curve (AUC) were employed to evaluate the accuracy of the models for classification. The observed results showed a better performance under SVMs than under artificial neural network and classification tree algorithms. Moreover, SVM with polynomial function (SVM-poly) worked slightly better than SVM with Gaussian radius basis function. The overall accuracy, kappa statistic, and AUC of SVM-poly were 0.943, 0.79, and 0.944, respectively. Meanwhile, the classification accuracy was 0.794 for clay, 0.992 for loam, and 0.661 for sand under SVM-poly. Elevation, terrain classification index for lowlands, and flow path length were the most important terrain indicators affecting the variation in the soil texture class in the study area. These results showed that the support vector machines are feasible and reliable in the identification of soil texture classes.

1. Introduction

Soil properties play critical roles and thus have influence in agricultural engineering, such as in soil improvement, land consolidation, management of drainage, soil erosion and irrigation. Knowledge on the variations in the soil properties could provide valuable information to design a more rational use and management plan especially in the cultivated areas. Climate, biota, and geological history are the critical factors affecting the chemical and physical properties of soil at larger scales (e.g., regional and continental scales), while human activities and topography may be the dominant factors controlling the properties of soil at smaller scales. Topography has significant impacts on runoff, drainage, and soil erosion and hence on the soil development (Jenny, 1941; Moore et al., 1991, 1993). The relationship between chemical and physical properties of soil and terrain indicators has been investigated intensively (Malo et al., 1974; Moore et al., 1993; Govers et al., 1996; Thompson et al., 2006; Wilcke et al., 2008; Wu et al., 2008; Leiß et al., 2012; Guo et al., 2013; Zhang et al., 2014). For example, Malo et al. (1974) found that soil clay, surface thickness, or organic C increased from shoulder position to foot slope along a hill slope. In southern Ecuadorian Andes, the sand/clay ratio of surface horizon increased with the increase of elevation and slope (Ließ et al., 2012). In southwestern China, tillage controlled the redistribution of soil particles of sloping terrace with embankment on a hill slope (Zhang et al., 2014).

Classical statistical methods have been applied to explore the relationship between terrain parameters and the chemical and physical properties of soil (Moore et al., 1993; Gessler et al., 2000; Gobin et al., 2001; Wilcke et al., 2008; Wu et al., 2008; Guo et al., 2011). Recently, several machine learning methods such as artificial neural networks, decision trees, and support vector machines have been proposed as alternative techniques for soil mapping (Zhao et al., 2009; Kovačević et al., 2010; Ehret, 2010; Guo et al., 2013; Leiß et al., 2012; Brungard et al., 2015; Taghizadeh-Mehrjardi et al., 2015, 2016; Zádorová et al., 2015; Heung et al., 2016). Much of the above focused on the classification of soil and the outcome of these investigations demonstrated that the performance of models varied under diverse circumstances (Ehret, 2010; Kovačević et al., 2010; Brungard et al., 2015; Taghizadeh-Mehrjardi et al., 2015; Heung et al., 2016). For example, using the ground penetrating radar data to classify rock layers, Ehret (2010) compared the differences between artificial neural networks and

* Corresponding author.
E-mail address: wuwei_star@163.com (W. Wu).

https://doi.org/10.1016/j.compag.2017.11.037
Received 9 January 2017; Received in revised form 20 November 2017; Accepted 26 November 2017
0168-1699/ © 2017 Elsevier B.V. All rights reserved.

support vector machines (SVMs). The results showed that the multi-class SVM presented better pattern recognition. In addition, Taghizadeh-Mehrjardi et al. (2015) compared six data mining classifiers, namely, logistic regression, artificial neural network, support vector machine, k-nearest neighbor, random forest, and decision tree model, to predict the spatial distribution of soil groups in Iran. They found that both the decision tree model and artificial neural network produced a higher accuracy and kappa index than others. Brungard et al. (2015) evaluated 11 machine learning classifiers for mapping the soil taxonomic classes in semi-arid areas and reported that the random forests using covariates selected via recursive feature elimination were consistently the most accurate or performed best. In digital soil mapping, for classification purposes, Heung et al. (2016) compared a variety of machine learners and found that the model choice and sample design greatly influenced the outputs.

Soil texture is an important physical property of soil which plays a key role in soil processes including water retention and soil fertility. Studies have demonstrated that the variations in the soil texture are closely related to topography (Gobin et al., 2001; Brown et al., 2004; Wilcke et al., 2008; Zhao et al., 2009; Leiß et al., 2012). For instance, Gobin et al. (2001) used stepwise multiple linear regression to predict the soil texture at the surface horizon based on terrain attributes over a catchment in southeastern Nigeria. They found that the smaller particle sizes (clay and silt) correlated well with the slope gradient and compound topographic index, whereas larger particle fractions correlated better with contributing area and stream power index. In the southern Ecuadorian Andes, Ließ et al. (2012) reported that the elevation was the most important variable controlling the soil texture of surface horizon based on the random forest method. Zhao et al. (2009) applied an artificial neural network to predict the soil texture over a watershed in Canada, based on a set of hydrographical parameters that were derived from a digital elevation model.

In this context, the objective of the current study is to evaluate the ability of the support vector machines, artificial neural networks, and decision tree classifiers for identifying the soil texture class based on terrain parameters. The relative importance of terrain parameters is then investigated by the best model. The observed results from this investigation provide useful information for mapping the categorical properties of soil such as soil type and soil texture.

2. Materials and methods

2.1. Study site

This study was conducted in a small watershed located in the core area of Three Gorges of the Yangtze River, southwest China (Fig. 1) having a moderate sub-tropical climate with a mean annual precipitation of 1224 mm and a mean temperature of 15 °C. The average annual sunlight is about 1477 h and the relative humidity is about 75%. The topography of the study region is mountains, where the elevation ranges from 238 to 1631 m and the major land is slope fields with the slopes between 0 and 63°. The annual crop rotation is rape (*Brassica napus* L.), corn (*Zea mays* L.) or sweet potato (*Ipomoea batatas* L.).

The parent materials of soil were developed from two soil strata (Xujiahe and Daye formations, Gong, 1999). These two formations were deposited during the late Triassic period and in the early Triassic period, respectively. The Xujiahe formation is composed of numerous types of rock, including glutenite, fine sandstone, siltstone, etc. The Daye formation is dominated by alternating mudstone, marls and thinly-bedded limestone. According to FAO soil classification, soils are classified as Regosols (when developed from Xujiahe formation) and Entisol (when developed from Daye formation) (FAO, 1988). In the plough layer, most of the field soils are loamy, clayey and sandy. Soil pH varies from 5.3 to 8.6 and the soil organic matter ranges from 6.3 to 26 g/kg.

2.2. Soil texture class

In September 2012, a total of 1032 soil samples at a depth of 20 cm were collected from the cultivated soils of slope fields. Ten sub-samples were randomly collected from each field and mixed as a composite sample. Soil texture class was estimated by twisting the composite sample between fingers according to the flowchart (Thien, 1979; Table 25 in FAO (2006)). This approach is referred as “texture-by-feel” or “Fingerprobe” (Sponagel et al., 2005), which is considered to be a suitable alternative to the laboratory measurement of soil texture which is generally time consuming and is not cost-effective (Foss et al., 1975; Post et al., 1986; Pachepsky et al., 2006; Vos et al., 2016). Soil texture could be divided into general or broad classes (3 or 4) or very fine classes (13 or 20). In practice, it is convenient to indicate the general classes (USA, soil survey manual, P110). For example, sand, loam, clay, and silt are given in FAO (2006), Fig. 4). The current work employs the general classes of soil texture. The following field criteria for estimating the soil texture were applied to fit the soils of the area (FAO, 2006, Table 25).

Sand: not possible to roll a wire of about 7 mm in diameter.

Loam: possible to roll a wire of about 3–7 mm in diameter, but breaks when trying to form the wire to a ring of about 2–3 cm in diameter, moderately cohesive, adheres to the fingers.

Clay: possible to roll a wire of about 3 mm in diameter and to form the wire to a ring of about 2–3 cm in diameter, cohesive, sticky, gnashes between teeth, and has a moderately shiny to completely shiny surface after squeezing between fingers.

In order to assess the estimates of field-based texture class, the particle sizes of 43 samples were analyzed using a pipet method. The corresponding soil texture classes were identified by using a soil texture calculator (https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid = nrcs142p2_054167). The overall accuracy and kappa index of general classes measured by the experienced soil scientists in the field and by the lab analyses were 76.7% and 0.604, respectively. Hence, the data were suitable for further analysis.

The total numbers of clay, loam, and sand soil samples for the two formations were 63, 854, and 115, respectively. The numbers of clay, loam, and sand soil samples were 11, 445, and 39 for the Xujiahe formation and 52, 409, and 76 for the Daye formation, respectively. Obviously, most of the samples with clay texture were collected from the Daye formation.

2.3. Terrain parameter

Topography plays a critical role in influencing the variations of the soil properties and a number of terrain indicators could be derived from digital elevation models (DEMs) (Moore et al., 1993; Florinsky et al., 2002). For instance, the variations of soil properties in the surface layer are either positively or negatively related to terrain indicators (Florinsky et al., 2002; Wu et al., 2008; Tajik et al., 2012; Mehnatkesh et al., 2013). Compared to the labor intensive as well as cost intensive measurements of soil properties, topographical data are easily obtained. However, some of the terrain indicators are highly correlated which may not precisely estimate the partial regression coefficient and also the relative importance of independent variables (Montgomery et al., 2001). Fortunately, the multicollinearity could be detected by the variance inflation factor (VIF), which can be calculated as follows (Marquardt, 1970),

$$VIF_i = \frac{1}{1-R_i^2} \quad (1)$$

where, R_i^2 is the coefficient of determination of a regression of the i^{th} independent variable on all of the other independent variables. The independent variables with VIFs that are less than or equal to 5 can be kept in the model (Montgomery et al., 2001).

In this study, forty-five terrain indicators were calculated by the

Download English Version:

https://daneshyari.com/en/article/6539859

Download Persian Version:

https://daneshyari.com/article/6539859

Daneshyari.com