Soil & Tillage Research 155 (2016) 19–26

Contents lists available at ScienceDirect

Soil & Tillage Research

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

A comprehensive support vector machine-based classification model for soil quality assessment

Yong Liu[a,*], Huifeng Wang[a], Hong Zhang[a], Karsten Liber[a,b]

[a] Institute of Loess Plateau, Shanxi University, Taiyuan, Shanxi, China
[b] Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form 11 July 2015
Accepted 14 July 2015

Keywords:
Soil quality assessment
Support vector machine
Comprehensive classification model
Heavy metal contamination
Soil fertility

ABSTRACT

Soil quality is defined here as the capacity of soil to have biological function, to sustain plant and animal production, to maintain or enhance water and air quality and to support human health and habitation. There are different soil quality assessment models based on diverse methods and data, but none of the models can fully meet all purposes. The selection of an appropriate soil classification model therefore becomes an important aspect in soil quality assessment. This paper presents a new comprehensive support vector machine-based classification model for classification of urban soil quality and then uses that model to assess the soil quality of Taiyuan relative to Chinese environmental quality standards and regional background values. The results indicated that the support vector machine-based soil quality model combined soil heavy metal contamination and soil fertility data satisfactorily, with an accuracy of 98.3333%. The soil quality of Taiyuan was subsequently divided into five classes (IA, IB, IC, IIA and IIB). Fifty percent of all samples were classified as class IB, indicating that soil quality within the study area was good. This paper shows that a comprehensive support vector machine-based classification model is feasible and reliable for soil quality assessment. Furthermore, the assessment presented could provide references for related ecological problems.

1. Introduction

With the rapid industrialization and urbanization that has occurred in most cities of the world in recent decades, there have been increasing concern for and attention on soil quality. The main threats to soils are identified as a decline in organic matter, increased soil erosion, compaction and salinization and an increased probability of floods, landslides, contamination, acidification and sealing (Montanarella and Rusco, 2008). Urban soil is an important component of urban ecosystems, and it provides a medium and nutrients to urban landscape plants and crops. It is also a sink and source of urban contamination and thus can affect urban eco-environmental quality and human health (Cachada et al., 2012; Praveena et al., 2014). Urban soil quality is influenced by both natural and anthropogenic factors, but humans play the most influential role in altering the performance characteristics of soils by mining, industry, agriculture, waste treatment and transportation (Karim et al., 2014; Yu et al., 2014).

The assessment of soil quality is generally based on a combination of soil environmental quality, soil productivity and soil sustainability. Urban soil environmental quality has often been impacted by a substantial amount of contaminations, including heavy metals, originating from different sources. Heavy metals do not degrade by biological or biochemical processes, although their speciation may change, become bioaccumulated and in a few cases, biomagnified along the food chain (Liao et al., 2006). Hence, heavy metals in soils are widely considered to potentially have unfavorable environmental impacts, as well as affect human health and prosperity (e.g., economic losses, safety of agricultural products and endangerment of human health, especially children) (Guo et al., 2012; Praveena et al., 2014). In many cases, urban soil productivity has been impacted by anthropogenic factors such as fertilization and wastewater irrigation, which increase soil fertility (Bhaduri and Purakayastha, 2014). Soil fertility in turn can affect vegetation composition, plant community biomass and plant functional group biomass. Many researchers and practicing farmers have also observed that fertile soil with high soil organic matter content and diversity of soil microbiota enhances vegetation health through various processes (Zheng et al., 2015). Urban soil also provides nutrients for both crops and ornamental plants. Thus, soil fertility is another important parameter to assess to determine urban soil productivity and sustainability.

* Corresponding author.
E-mail address: Liuyong@sxu.edu.cn (Y. Liu).

http://dx.doi.org/10.1016/j.still.2015.07.006

The risk assessment methods used to evaluate soil heavy metal contamination and soil fertility assessment are diverse (Bhuiyan et al., 2010; Li et al., 2014a; Saby et al., 2009). Traditional methods of soil heavy metal contamination assessment include the single-factor index and the integrated pollution index (Xu and Hu, 2014), the analytic hierarchy process (AHP) (Chen et al., 2012), the fuzzy comprehensive evaluation method (Fu et al., 2014; Hui, 2013), and the geo-accumulation index method (Akbulut et al., 2013). The geo-accumulation index method is recommended by the U.S. Environmental Protection Agency and widely used by scholars. The traditional methods of soil fertility assessment include the analytic hierarchy process (AHP) (Zhou et al., 2009), the fuzzy comprehensive evaluation method (Fang, 2012) and the gray correlation analysis method (Zhang et al., 2009). While many researchers are concerned with the assessment of soil heavy metal contamination or soil fertility (Taylor et al., 2010; Teng et al., 2014), it is rare that research combines two approaches to assess soil quality. This is possibly because there is a complex nonlinear relationship between soil heavy metal content and soil fertility and because traditional methods do not perform well in addressing with the complex nonlinearity. Additionally, because the weights in assessment indices are artificially set in soil assessments, the results obtained by traditional methods often lack reliability, objectivity and currency (Jiang et al., 2014).

Machine learning algorithms, such as artificial neural networks (ANNs), *k*-means algorithms, genetic algorithms, decision trees, support vector machines (SVMs) and multiple linear regression methods have been used to advance classification and forecasting in recent years. Among these methods, ANNs have the most complex mathematical structure and can simulate human learning and pattern recognition (Were et al., 2015). However, due to the lack of theoretical results from a statistical perspective, as well as the low interpretability of this class of black-box models, some alternative strategies have been considered (Poggi and Portier, 2011).

As a more robust statistical method, SVMs are now widely used for environmental assessment with satisfactory performance (Aryafar et al., 2012; Jiang et al., 2014). The SVM method is an artificial intelligence machine learning theory introduced in the early 1990s as a non-linear solution for regression and classification tasks (Behzad et al., 2009; Vapnik, 1995). It relies on the statistical learning theory or VC theory, which enables learning machines to find important support vector information using a very small number of parameters (Bishop, 2006; Kovačević et al., 2010; Wang, 2005). Many classification algorithms are based on an independence assumption and are thus greatly influenced by the correlation among characteristics, but SVMs are not sensitive to this. SVMs obey the structural risk minimization principle (SRMP), which has been shown to be superior to many other modeling techniques obeying the traditional Empirical Risk Minimization Principle (ERMP) (Araghinejad, 2014). This technique has been proven to have superior performances in addressing various problems due to its generalization abilities, robustness against noise and interferences (Steinwart and Christmann, 2008) and its computational efficiency compared with several other methods, such as neural networks and fuzzy networks (Wang, 2005; Were et al., 2015). The published literature has shown that although the SVM method has been used to address many environmental problems (Jiang et al., 2014; Aryafar et al., 2012), it has rarely been used in research on soil quality assessment.

The aims of this study were (1) to develop a comprehensive method to assess soil quality; (2) to explore the application of the SVM method in the classification of heavy metal levels in soils; (3) to assess soil quality by using the SVM method in combination with the levels of soil heavy metal contamination and soil fertility; (4) to identify the distribution of heavy metals and the pattern in soil quality in Taiyuan city; and (5) to describe correlation among metals and other parameters in soils.

Therefore, in this paper, we present a comprehensive soil quality assessment model based on SVM methodology for Taiyuan city, which could be applied in other areas when evaluating soil heavy metal contamination and soil fertility, etc. Furthermore, this paper provides useful information regarding soil quality and soil management in Taiyuan.

2. Materials and methods

2.1. Area description

Taiyuan city is located on the east edge of the Loess plateau and at the center of Shanxi Province with heavy industries (Fig. 1). It is heavily industrialized with structurally complex factories concentrated in specific areas. Large volumes of waste products and emissions are discharged to the local environment and have resulted in the accumulation of heavy metals in soils, especially soils in the suburbs. Previous studies have shown that the heavy metals Hg, Cd and Pb have been enriched in the surface soil of Taiyuan (Li et al., 2004). In addition, soil contamination in the

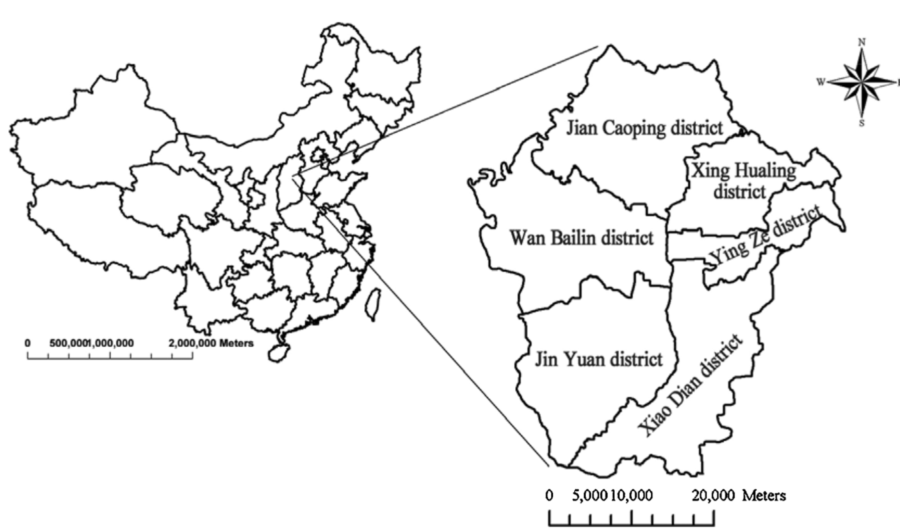

Fig. 1. Location of research area.