

Use of Data Mining Techniques to obtain rules and standards regarding Pedotransfer Functions developed for Soil Water Retention

Marcel P.V. Veloso (IC), Luiz Henrique A. Rodrigues (PQ)

Abstract

Pedotransfer functions (PTF) are developed in order to determine some soil parameters whose measurements are laborious, expensive and time demanding. Besides that, they should not be applied for a wide range of soils due to errors that it might result for a large scope of soils other than those for which it was developed. The purpose of this research is to use the Rule Induction of Data Mining Techniques to increase the reliability, accuracy, coverage and scope of the PTF.

Key words: Soil Water Retention, Data Mining, Rule Induction

Introduction

Pedotransfer functions (PTF) are mathematical models developed to use soil parameters to calculate others. According to Bouma (1989), they aim to "transform data we have in data we need".

Data Mining, a computational resource and information associated with the achievement of patterns, is shown to be a useful tool for determining errors of PTF.

Rule Induction is one of the most important Data Mining techniques, since "hidden" patterns in the database are often expressed as rules. This method is one of the most important to discover knowledge in databases. (GRZYMALA-BUSSE et al., 2005).

Results and Discussion

Based on the qualitative analysis of the rule evaluator, most of the rules showed accuracy above 65%, which is a good example of how the rules are useful to determine many hidden patterns of PTF errors.

Among all developed results, one of the most important patterns was the rule obtained for Tomasella et al. (2003), which could not be visualized by Veloso and Rodrigues (2013) as shown in Image 1.

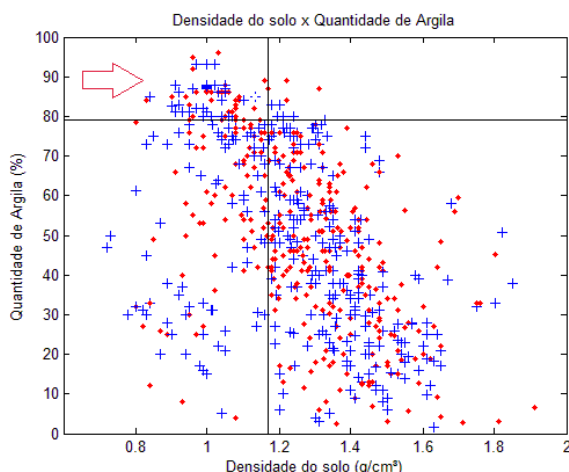


Image 1. Example of the results and association obtained by Veloso and Rodrigues (2013).

According to Image 1, it is not possible to visually find patterns considering Wilting Point conditions for Tomasella et al. (2003) (Veloso and Rodrigues, 2013). However, according to the red arrow shown in the image 1, the Data Mining technique made possible the achievement of the rule regarding the understanding and evaluation of hidden error patterns, which was previously unidentified in that area.

Hence, the pattern obtained by the Rule Induction technique (algorithm JRip) associated different levels of clay, bulk density and carbon amount with the probability of occurring an error in the PTF evaluated. In other words, for a clay amount higher or equal to 79%, bulk density less than or equal to 1.17 g / cm³ and carbon amount higher than or equal to 0.66% but less than or equal to 1.4%, the error generated by the PTF would be categorized as acceptable. If not categorized as acceptable, that will be an essential point to be eventually focused at when studying soils with distinct characteristics.

Conclusions

Therefore, by using Data Mining Rule Induction technique, it was possible to achieve a deep knowledge regarding the errors associated with the PTFs evaluated. Knowledge that was not possible to be visually identified and easily studied.

Acknowledgement

I would like to acknowledge and thank my sponsor Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and University of Campinas (Unicamp) for letting me achieve my potential to research.

BOUMA, J. *Using soil survey data for quantitative land evaluation. Advances in Soil Science*, v.9, p.177-213, 1989.

GRZYMALA-BUSSE, J.W.; *Rule induction. In Data Mining and Knowledge Discovery Handbook*, p 277-294, 2005.

TOMASELLA, J; PACHEPSKY, Ya.; CRESTANA, S.; RAWLS, W.J. *Comparison of Two Techniques to Develop Pedotransfer Functions for Water Retention. Soil Sci. Soc. Am. J.*, v.67 p.1085 – 1092, 2003.

VELOSO, M.P.V.; RODRIGUES, L.H.A. *Identificação visual de padrões de retenção de água no solo. XXI Congresso de Iniciação Científica da Unicamp; 23-25 de outubro, Campinas; p.4-19, 2013.*