Geoderma 166 (2011) 15–33

Contents lists available at ScienceDirect

Geoderma

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

Application of fuzzy logic to Boolean models for digital soil assessment

J.J. de Gruijter [a,*], D.J.J. Walvoort [a], G. Bragato [b]

[a] Alterra, Wageningen University and Research Centre, PO Box 32, NL-6700 AA Wageningen, The Netherlands
[b] CRA – Centro per lo Studio delle Relazioni Pianta-Suolo, Via Trieste 23, 34170 Gorizia, Italy

ARTICLE INFO

Article history:
Received 21 January 2011
Received in revised form 1 June 2011
Accepted 4 June 2011
Available online 25 August 2011

Keywords:
Fuzzy sets
Fuzzy connective
Land evaluation
Grass farming
Seed-potato
de Morgan's law

ABSTRACT

Boolean models based on expert knowledge are often used to classify soils into a limited number of classes of a difficult-to-measure soil attribute. Although the primary data used for these classifications contain information on whether the soil is a typical class member or a boundary case between two classes, this is not retained in the final result. Such information is relevant in land use planning and soil management as it enables more flexible decision taking, but in the pre-digital era it was unfeasible to prevent the loss of it. We can now retain this information by fuzzifying the Boolean model using fuzzy logic. Choices must then be made on the type of membership function, logical operators, and formulation of the assessment rules. From a review of the main types of membership functions we conclude that piecewise linear functions are most appropriate in practical applications. Combinations of different fuzzy union (**or**) and intersection (**and**) connectives were tested on a 2-dimensional example. Nearly all combinations gave results that partly contradict the associated a priori knowledge, the exception being the Bounded sum connective for **or**, and the Product connective for **and**. We also found that in formulating the rules, overlap of predictor classes and negation should be avoided. Unrestricted choice of fuzzy connectives and rule formulation will generally lead to inconsistencies. The selected methods were tested in two case studies: one on suitability for seed-potatoes in an Italian region and one on suitability for grass farming in a Dutch region. The maps produced with the fuzzy and Boolean models are broadly similar. However, maps from the fuzzy models indicate that some areas represent a transition between two original Boolean classes, thereby providing relevant additional information. In the case study on seed-potatoes the quantitative prediction errors of the original Boolean suitability map were greatly reduced by the fuzzification.

1. Introduction

The aim of this study is to investigate how fuzzy logic can be applied to transform, or 'fuzzify', Boolean soil assessment models into models that predict assessment classes with gradual instead of abrupt transitions. First we discuss the concepts of digital soil assessment and fuzzy logic. Then we indicate some application aspects of fuzzification.

Digital soil assessment (DSA) has been defined by Carré et al. (2007) as "… the quantitative modeling of difficult-to-measure soil attributes, necessary for assessing threats to soil (erosion, decline of organic matter, compaction, salinisation, landslides, sealing, floods, and decline of biodiversity) and soil functions (biomass production, environmental interactions, physical support, production of raw material, cultural heritage, carbon pool, source of biodiversity (European Commission, 2006)), using DSM outputs", where DSM stands for digital soil mapping. See Carré et al. (2007) for a detailed analysis of the role of DSA in view of information needs in ecological risk assessment and policy guidance. We will refer to the soil attribute to be assessed as the 'target variable'. This may be qualitative, defined at a nominal or ordinal scale, or quantitative, defined at an interval or ratio scale.

The classical and ubiquitous example of DSA is *land evaluation*, where a digitized soil map in a GIS is transformed into a suitability map. These transformations are defined by a land evaluation model that predicts a suitability class for each of the map units of the soil map. Many other forms of DSA have been developed in recent years, for instance, digital mapping of soil quality, vulnerability for erosion and potential for nitrate leaching.

Broadly speaking, three different types of models are used in DSA:

1. Statistical models, for instance linear regression models, regression trees and neural networks. These models are primarily based on measurements and represent empirical rather than causal relationships. This is why they are often referred to as 'black-box' models. Their applicability for DSA is limited by lack of data, as the target variables to be predicted are by definition difficult to measure.
2. Mathematical process models, for instance simulation models. These models are based on quantitative theory of the underlying

* Corresponding author. Tel.: +31 317 474238; fax: +31 317 419000.
E-mail address: jaap.degruijter@wur.nl (J.J. de Gruijter).

0016-7061/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2011.06.003

processes. As opposed to statistical models, they represent causal relationships. For this reason they may be referred to as 'white-box' models. Their applicability for DSA is limited too, by lack of precise knowledge of the underlying processes or by lack of data as some models require many data to calibrate their parameters.

3. Models based on expert knowledge, for instance land evaluation models. The available expert knowledge stems typically from practical experience. Although often laid down in a clear-cut assessment table or tree, this knowledge is qualitative and, to a certain extent, imprecise and vague. In this case we can speak of 'gray-box' models (Lindskog, 1997). They are frequently the only practicable option for DSA, and this motivated us for this study.

We assume that a Boolean DSA model already exists, and so the most important work, i.e. expressing of expert knowledge in a table or tree, has already been done. See Kaufmann et al. (2009) for a study on fuzzy DSA where expressing of expert knowledge was part of the problem. The Boolean model may have been used in practice, or it may have been constructed as an intermediate step in the development of a fuzzy DSA model. In either case, what rests is to fuzzify the Boolean model. Although this is a limited and mainly technical task, it appears not as straightforward as it may seem at first glance.

The structure of this paper is as follows. Section 2 describes what in this study is the source material: Boolean assessment tables and trees, and the derivation of a Boolean rule base from these.

Section 3 summarizes the theory of fuzzy sets, fuzzy logic and fuzzy models, as far as relevant to soil assessment. The theory of fuzzy sets is an extension of classical set theory. Where classical set theory starts from the pre-assumption that there are only two possibilities – an element does or does not belong to a set – fuzzy set theory allows that an element belongs to a certain degree to a set or class. In principle, the latter approach is more in line with the fact that boundary cases are abundant in soil assessment, and that the crisp class boundaries as normally used at present are often more or less arbitrary. Fuzzy logic operates with fuzzy sets and is the fuzzy analog of Boolean logic. With methods based on fuzzy logic the arbitrariness or vagueness of class boundaries can be accounted for, by creating fuzzy instead of crisp boundaries. This leads to a more intensive use of available data and knowledge, and to more differentiated assessments.

Section 4 discusses how fuzzy logic can be applied in soil assessment, focusing on the type of membership function and on the logical operators. In earlier studies modeling of interactions (i.e. the effect of a given primary variable on the target variable depends on the level of one or more other variables) is a weak point. Therefore, the present study pays special attention to that issue. Section 5 discusses briefly how the results can be presented cartographically.

In Sections 6 and **7** two case studies are presented. The first is on suitability for seed-potato in Calabria, Italy. This case is relatively simple, as the suitability is directly inferred from the primary soil variables by a one-stage procedure. The second case is on suitability for grass in Noord-Brabant, The Netherlands. This is more complicated, as the inference follows a two-stage procedure. First, suitability factors are predicted from the primary variables. Second, suitability is predicted from these factors. Section 7 ends with conclusions.

2. Boolean assessment tables, trees and rule bases

2.1. Introduction

As a preparation to Boolean assessment modeling, we first recapitulate some basic concepts from classical set theory and logic. The central concepts in classical set theory, developed by Georg Cantor at the end of the 19th century, are:

element the smallest entity that is taken into consideration in the context of a given application. This can be anything, for instance, a number, a soil unit, or a point in an area. The generic symbol for an element is x.

universe (also: universe of discourse) the collection of all elements that are relevant in the context of an application, for instance all real numbers between 0 and 100, all soil individuals in a given area, or all points in a given plane. A universe can be finite or infinite. The generic symbol for a universe is X.

set a group of elements from the universe considered. A set is often created in order to make a statement about all elements in it, or to apply a common treatment to them. A direct way to define a set is by listing all the elements that belong to it. An indirect way is to specify one or more properties that an element must possess in order to belong to the set. In that case one often speaks of a class. For instance: all soil individuals in a given area with a loam content in the topsoil greater than 10%. The generic symbol for a classical set is A.

Classical set theory is based on the 'law of the excluded middle', i.e. the principle that an element does or does not belong to a given set, and that these two cannot be true at the same time. Thus, according to this principle, for the membership of an element x in a set A there are only two mutually exclusive possibilities: member or not member. If, for instance, x represents a soil, and A is the set (class) of soils that are 'very suitable' for a given use, then this soil either does or does not belong to the class 'very suitable'. In mathematical notation: $x \in A$ or else $x \notin A$.

Similarly, classical logic pre-assumes that for any statement there are only two possibilities: 'true' or 'false'. In the example above: the statement 'soil x is very suitable' can only be true or false. Other possibilities than these two fall outside the framework of classical two-valued logic.

There are two rather common situations that seem to contradict this duality principle, but in fact agree with it. The first is that an element can belong to more than one set. For instance, if the class 'very suitable' is defined as a sub-class of the class 'suitable', then any soil in the class 'very suitable' belongs also to the class 'suitable', but not necessarily vice versa. This is an example of 'overlapping' sets. Indeed there are more than two possibilities here, namely: (1) 'suitable' and even 'very suitable', (2) 'suitable' but not 'very suitable', and (3) 'very suitable' neither 'suitable'. Here too, however, for the membership in any particular class there are still only two possibilities.

The second situation is that there are insufficient data available on the element to be classified, and therefore one is uncertain whether that element belongs to the class. Here too, there seem to be more than two possibilities, namely: yes, no and uncertain. However, uncertainty is a property of the person (the subject) who evaluates, not of the reality being evaluated (the object). Regarding the latter, classical set theory still considers only two possibilities: 'member' and 'not member'. One can only be uncertain about which of these two is true, and this uncertainty can be reduced or removed by collecting more or better data.

2.2. Boolean assessment modeling

Expert knowledge used for assessment is usually laid down in a table, with classes of predictors in the margins and assessment in the cells. For a simple example, suppose that a target variable '*suitability*' is to be assessed as 'good' or 'bad', on the basis of clay content (*cl*) and organic matter content (*om*). According to the experts, *suitability* is '*good*' if *cl* and *om* are both low (e.g. less than 5%), and '*bad*' in all other cases. This model is represented in as an assessment table in Table 1, and as an assessment tree in Fig. 1. An equivalent tree can be obtained by interchanging clay content and organic matter content.

In case of more than two predictors the table will be three- or multi-dimensional and the tree is more intricate, possibly divided into sub-trees. See Fig. 18 for an example of a tree with four predictors.