

LCC Decision tree analysis using ID3

N.M.Tamboli

Assistant Professor
Department of Agricultural
Engineering, Marathwada
Agricultural University,
Parbhani,
Maharashtra, 431402

A.M.Kamble

Assistant Professor
Department of Agricultural
Engineering, Marathwada
Agricultural University,
Parbhani,
Maharashtra, 431402

P.S.Metkewar

Associate Professor
Symbiosis Institute of
Computer Studies and
Research (SICSR), Atur
Centre, Deep Bunglow Chowk,
Model Colony, Pune 411 016

ABSTRACT

The basic principle of soil and water conservation is to use the land according to its capability. Therefore the knowledge of land capability classification is a prerequisite and important for planning of watershed development programme. The capability of land depends upon the characteristics of the land like slope, soil depth and soil type. Considering this input parameters Land Capability Classification (LCC) decision tree is formulated using ID3 algorithm. Results shown by generated decision tree are compared with the results obtained theoretically by considering ranges given in the standard table and were found similar. After successful testing of decision tree it is concluded that the model is useful for suggesting the soil and water conservation practices for watershed development plan.

General Terms

Decision Tree, Application of ID3 algorithm.

Keywords

Land Capability Classification (LCC), Model, Watershed, Land, ID3, Decision Tree.

1. INTRODUCTION

Land refers to a delineated part of the earth's surface, including mountains, rivers, swamps, deserts, islands, and coast areas. Soil comprising minerals and biological materials is a natural feature of the earth's surface and support plant growth. Land characteristics are influenced by the atmosphere, the soil and underlying geology, the relief, the hydrology, the plant and animal populations and finally the human activities. Any land area for the purpose of being utilized for agricultural production not only need certain initial modifications but also appropriate maintenance later. The land might need initial clearing, leveling and proper field layout. It might also need protection from the erosive actions of wind and water. Harmful substances or excess water has to be removed from the land area. This all can be achieved with the watershed management in relation to land use and land capability classification. The basic principle of soil and water conservation is to use the land according to its capability and treat the land according to its need. Therefore the knowledge of land capability classification is a prerequisite and important for planning of watershed development programme. The present study was undertaken with the objective to develop decision tree for land capability classification (LCC).

2. PROBLEM DEFINITION

Determination of land capability class is a complex and technical process. To simplify this process for watershed planner, software tools can be used effectively. Use of advanced technology for precise planning of watershed development is today's need [1], therefore the present study was undertaken with the objective to develop a decision tree for finding the land capability class which in turn will help to develop decision support system to suggest soil and water conservation practices for watersheds.

3. MATERIALS AND METHODS

For building decision tree training data set is extracted from the referenced literature [2][3].

3.1 Basic Factors

3.1.1 Slope

Various ranges of slopes for land capability classes are given in the table number 1.

Table 1. Ranges of Slopes for Land Capability Classes

Notation	Slope class	Land capability class	Slope ranges (%)
A)	Nearly level	Class I	0-1%
B)	Gentle sloping	Class II	1-3%
C)	Moderate sloping	Class III	3-5%
D)	Strongly sloping	Class IV	5-10%
A)	Nearly level	Class V	0-1%
E)	Moderate steep	Class VI	10-15%
F)	Steep	Class VII	15-25%
G)	Very steep	Class VIII	>25

3.1.2 Soil Depth

The various ranges of soil depth for land capability classification are given in the table number.2.

Table 2. Ranges of Soil Depth

Symbol	Name	Depth range (cm)
D ₁	Very shallow	0-7.5
D ₂	Shallow	7.5-22.5
D ₃	Moderately deep	22.5-45.0
D ₄	Deep	45.0-90.0
D ₅	Very deep	>90.0

3.1.3 Soil Type

The various notations used for different soil types are given in the table number 3.

4. CONSTRUCTION OF DECISION TREE USING ID3 ALGORITHM

For assumed training data, ID3 algorithm generates a decision tree (popular classifier). The decision tree is generated on the basis of calculated Entropy and Information gain of assumed training data[4]. Using decision tree we can easily generate optimal tree in order to predict classification of new (unseen) training data.

The training data set used for building decision tree is shown in Table-4.

Table 4. Training data set used to build decision tree [2][3].

Texture	Depth	Slope	Class
S	d5	nearly_level	I
S-C	d5	nearly_level	I
S-L	d5	nearly_level	I
L	d5	nearly_level	I
C-L	d5	nearly_level	I
S-C-L	d5	nearly_level	I
S	d4	Gentle	II
S-C	d4	Gentle	II
S-L	d4	Gentle	II
C-L	d4	Gentle	II
S-C-L	d4	Gentle	II
S-C-L	d3	Moderate	III
L-S	d3	Moderate	III
S-C	d3	Moderate	III
C	d3	Moderate	III
C	d2	Strong	IV
S	d2	Strong	IV
R	d5	nearly_level	V
ST	d5	nearly_level	V
BS	d5	steep	VI
ST	d1	steep	VI
ES	d1	very_step	VII
R	d1	very_step	VII
BS	d1	very_step	VII

There are VIII classes with distribution shown in table 4.

I	II	III	IV	V	VI	VII	VIII
6	5	4	2	2	2	3	3

Entropy of a root node containing the whole training set as its subset is calculated[4]. $Entropy(rootNode.subset) = 2.88$

To identify the attribute to start the decision tree with, Calculations for Information gain for texture attribute are shown in Table 5. Similarly, the information gain for slope and depth is calculated[4].

$Gain(S, Texture) = 1.68$ $Gain(S, Depth) = 2.08$ **$Gain(S, Slope) = 2.64$**

Information gain for slope attribute is highest, hence the decision tree starts with slope attribute. Apply ID3 to each child node of this root, until leaf node or node that has entropy=0 are reached. Hence we have the child nodes as described below.

Slope attribute value nearly_level requires further analysis (application of ID3) to expand the tree. Following the above steps, the $Gain(S_{nearly_level}, Depth)$ and $Gain(S_{nearly_level}, Texture)$ needs to be calculated further.

Table 3. Notations for Soil Type.

Soil Texture	Character Notation used in training data set	Numerical Notation
Loam	L	1
Sandy Loam	S-L	2
Silty Loam	S-L	3
Clay Loam	C-L	4
Sandy Clay Loam	S-C-L	5
Silty Clay Loam	Si-C-L	6
Sandy Clay	SC	7
Silty Clay	SiC	8
Loamy Sand	L-S	9
Clay	C	10
Sandy	S	11
Rocky	R	12
Stony	S	13
Very Stony	VS	14
Boulder strewn	BS	15
Extremely stony	ES	16
Swampy	SW	17
Marshy	MA	18

$Gain((S_{nearly_level}, Depth) = 0.811278$

$Gain(S_{nearly_level}, Texture) = 3.811278$

Out of depth and texture attribute, information gain for texture attribute is more, hence texture is selected for next split. From table 7 it is clear that each resultant node is leaf node and hence no further splits are possible.

Based on table 6 & 7 the results were tested using open source data mining tool WEKA [5] and Dataminer [20] , which were found similar.

Table 5. Entropy and Information gain calculations for texture attribute.

Count	1	2	3	4	5	6	7	Subset Entropy	Weighted Sum
Texture									
S (3)	1	1	0	1	0	0	0	1.584963	0.19812
S-C (3)	1	1	1	0	0	0	0	1.584963	0.19812
S-L (2)	1	1	0	0	0	0	0	1	0.083333
L (1)	1	0	0	0	0	0	0	0	0
C-L (2)	1	1	0	0	0	0	0	1	0.083333
S-C-L (3)	1	1	1	0	0	0	0	1.584963	0.19812
L-S (1)	0	0	1	0	0	0	0	0	0
C (2)	0	0	1	1	0	0	0	1	0.083333
R (2)	0	0	0	0	1	0	1	1	0.083333
ST (2)	0	0	0	0	1	1	0	1	0.083333
BS (2)	0	0	0	0	1	0	1	1	0.083333
ES (1)	0	0	0	0	0	0	1	0	0
				Information Gain (Texture)					1.094361

Table 6 : Child nodes and information gain for each child node

Child Node	Information Gain
nearly_level	0.27 (Needs further analysis)
Gentle	0 (Decision : LCC - II)
Moderate	0 (Decision : LCC - III)
Steep	0 (Decision : LCC - VI)
Strong	0 (Decision : LCC - IV)
very_step	0 (Decision : LCC - VII)

Table 7 : Child nodes and information gain for each child node

Texture	Information Gain
S	0 (Decision : LCC - I)
S-C	0 (Decision : LCC - I)
S-L	0 (Decision : LCC - I)
L	0 (Decision : LCC - I)
C-L	0 (Decision : LCC - V)
S-C-L	0 (Decision : LCC - V)

5. RESULTS AND DISCUSSION

The developed decision tree using ID3 algorithm is tested with different conditions and the results are noted. Further the generated decision tree was validated using the open source data mining tool Weka [5], the generated decision tree is shown in figure 1.

```

slope = nearly_level
| Texture = S: I
| Texture = S-C: I
| Texture = S-L: I
| Texture = L: I
| Texture = C-L: I
| Texture = S-C-L: I
| Texture = L-S: null
| Texture = C: null
| Texture = R: V
| Texture = ST: V
| Texture = BS: null
| Texture = ES: null
slope = Gentle: II
slope = Moderate: III
slope = Strong: IV
slope = steep: VI
slope = very_step: VII

```

Figure 1 : Resultant Tree Diagram [5]

Thus the following conclusions are obtained from this study.

1. Results generated by the LCC decision tree are correct and similar to the results obtained theoretically by considering ranges given in the standard table.
2. After successful testing it is concluded that this decision tree could be useful for identification the land capability class for watershed management plan.

3. After successful testing it is concluded that the generated model (machine learning technique) is useful for developing decision support system for suggesting the soil and water conservation practices.

6. FUTURE SCOPE

For validation purpose, the training data set was processed using Dataminer [13] Classifier software tool, if information ratio method is applied in place of Information gain the resultant tree generated indicates that depth attribute can be selected as start point instead of slope. This generated model is to be validated using sample field data.

7. ACKNOWLEDGMENTS

We acknowledge the developers of Weka and Dataminer tools which were used for testing the results.

8. REFERENCES

- [1] Sidhu, G.S.; Rana K. P.C.; Lal Taarsem; Mahapatra, S.K.; Verma, T.P.; Rao, R.V.S. (2007). Land capability classification and assessment of soil degradation status for suggested land use in soils of Himachal Pradesh J. Indian. Soc. Soil. Sci. 55 (3).
- [2] Tidemand, E.M. (2000). Land capability classification for watershed development : 38-60.
- [3] Tejuwani, K.G. Dhruvanarayana (1961). Soil conservation survey and land use capability planning in the Ravine lands of Gujarat J. Indian Soc. Soil. Sci., 9 : 233-34.
- [4] Dunham Margaret H and S. Sridhar (2007). Data Mining Introductory and Advanced Topics ,Pearson Education,ISBN 81-7758-785-4.
- [5] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
- [6] Shrinivasan, T.R. and Kamphorst, A. (1971). Placement of medium and deep cracking clay soils in the comprehensive system of soil classification J. Indian. Soc. Soil. Sci. 19 : 415-421.
- [7] Tejuwani, K.G. and Das D.C. (1977). F.A.O. Soils bulletin, 33p 11.
- [8] Alok Kumar and Tripathi, R.P. (1987). Land Use pattern and some soil physical properties of mini watershed. J. Indian Soc. Soil. Sci. 2 : 262-67.
- [9] <http://www.doc.ic.ac.uk/~sgc/teaching/v231/lecture11.html>
- [10] http://www.en.wikipedia.org/wiki/ID3_algorithm
- [11] <http://www.cs.waikato.ac.nz/ml/weka/>
- [12] <http://rapid-i.com/>