Predicting the comercial quality of slate slabs with a mathematical model

Carla Iglesias*, Javier Martínez**, Javier Taboada*, Eduardo Giráldez*

*Department of Environmental Engineering. University of Vigo, Spain.
Corresponding author: carlaiglesias@uvigo.es
**Centro Universitario de la Defensa de Marín. Marín, Pontevedra, Spain

Abstract

Slate is a natural material whose impermeability, fisibility and resistance to climate make it a valuable natural resource for construction purposes. As a natural geological material, it can be affected by some singularities that alter its mechanical properties as well as the aesthetic. These singularities have always been assessed manually by an expert, who classifies each slate into a certain commercial quality.

The purpose of our research work is to build a model able to classify slate slabs according to their quality from the inspection of images obtained from a computer vision system. A set of 70 slabs was used, knowing the commercial quality assigned by two experts. The computer vision system acquires the range, grayscale and RGB images from each slab, then some algorithms are used to inspect these images and a vector of characteristics is obtained. These characteristics will be used to predict the commercial quality (first quality, second quality or reject).

This classification problem was approached with the following models: Classification and Regression Trees, Multilayer Perceptron and Support Vector Machines. The best results correspond to the first model (CART), while neither MLP nor SVM were able to reproduce a proper classification criterion. This fact points out the linear separability of the classes of this classification problem.

Key words: Classification and Regression Tree, Multilayer Perceptron, Support Vector Machines, Slate Quality Classification

1 Introduction

Slate, one of the most employed natural materials for construction purposes, is mostly used for roofing, installing thin slabs of different geometries. After extracting irregular slate blocks from the quarry, they go through a manufacturing process that involves many cutting stages until the final slabs are obtained. As a natural geological material, slate can display some singularities that can affect not also its aesthetic but its mechanical properties. These singularities have always been assessed manually by an expert, who classifies each slate into a certain commercial quality after

taking a look at it. In order to achieve an optimum, objective classification, an automatic process is pursued by means of mathematical techniques.
This research work developes an automatic classification system that comprises artificial vision techniques and mathematical algorithms to inspect the images, extract the valuable information and determine the commercial class quality to each sample. The work started in [2] is continued, where a different vision system (not including color information) gathered the images from slate slabs and a new DAGSVM algorithm was used with the same aim as in this case. That multiclassification algorithm, based on SVM binary classifiers with a one-versus-all approach, was presented in Mathematical Modelling in Engineering & Human Behaviour 2013 conference.
A set of 70 slabs was used, knowing the commercial quality assigned by two experts. The computer vision system acquires the range, grayscale and RGB images from each slab, then some algorithms are used to inspect these images and a vector of characteristics is obtained. The singularities considered are: surface uniformity, kink-bands, false squaring, warping, material defects, presence of carbonates and presence of sulfides. These characteristics will be used to predict the commercial quality (first quality, second quality or reject).

2 Results and discussion

After the inspection of the range, grayscale and RGB images of the slabs, the variables shown in Table 1 were defined. These variables are the input data for the model, and two cases were considered: first, *Input A* was defined as the set of main variables; then, *Input A* was defined as the main variables together with the auxiliary variables (marked with (+) in Table 1).
With respect to the output of the model, it is the commercial quality class defined by two experts: Expert 1 and Expert 2. They considered three different cases: first quality (label 1), second quality (label 2) and reject (non-markeatable slab, label 3), depending on the singularities observed.
This classification problem was approached with the following models: Classification and Regression Trees [1], Multilayer Perceptron [3] and Support Vector Machines [4]. The results of the predictions are summarized in Table 2.
According to Table 2, CART is clearly the technique which predicts the best the commercial quality of this 70-slab dataset. Both MLP and SVM are unable to replicate the criterion of the expert, and eventually assign the most common class in the dataset. Criterion of Expert 1 is better predicted than that of Expert 2, and a minimum test error rate of 0.057 (5.7% is achieved with the tree showed in Figure 1). Whether Input A or B is considered, the optimum tree has the same architecture, meaning that the critical variables taken into account to classify each observation are included both in Input A and Input B. In other words, none of the auxiliar variables provide valuable information.

Table 1
Summary of the variables defined for each singularity.

Singularity	**Variable**	**Description**
Material defects	RatiosMD*$\in \mathbb{R}^4$	Ratios between real and theoretical areas
	RectangMD*$\in \mathbb{R}^4$(+)	Rectangularity factor
False squaring	AngleFE$\in \mathbb{R}^4$	Differences of corner angles with resp. to 90^o
	RectangFE*(+)	Rectangularity factor
Warping	AngleW$\in \mathbb{R}^{10}$	Differences between mean range values of corners and with respect to center
Surface irregularities	DevS	Deviation of pixel values
	NumEdges*$\in \mathbb{R}^4$	Number of edges
	MaxAmplEdg*$\in \mathbb{R}^4$	Maximum amplitude of edges
	MeanAmplEdg*$\in \mathbb{R}^4$	Mean amplitude of edges
	VarAmplEdg*$\in \mathbb{R}^4$	Amplitude variance of edges
	MeanDistEdg*$\in \mathbb{R}^4$(+)	Mean distance between edges
Presence of carbonates	MAreaCBW**$\in \mathbb{R}^3$	Mean area of int. reg. grayscale image
	MCircCBW**$\in \mathbb{R}^3$	Mean circularity of int. reg. grayscale image
	MRoundCBW**$\in \mathbb{R}^3$(+)	Mean roundness of int. reg. grayscale image
	MAreaCRGB**$\in \mathbb{R}^3$	Mean area of int. reg. color image
	MCircCRGB**$\in \mathbb{R}^3$	Mean circularity of int. reg. color image
	MRoundCRGB**$\in \mathbb{R}^3$(+)	Mean roundness of int. reg. color image
Presence of sulfides	MAreaS**$\in \mathbb{R}^3$	Mean area of int. reg. color image
	MCircS**$\in \mathbb{R}^3$	Mean circularity of int. reg. color image
	MRoundS**$\in \mathbb{R}^3$(+)	Mean roundness of int. reg. color image

** Vertical, horizontal and the two principal diagonals were considered.*

*** The study is performed by thirds.*

(+) Auxiliary variable.

Table 2
Summary of the results of CART, MLP and SVM models for the prediction of slate slabs commercial quality.

Input	Expert	CART	MLP	SVM
Input A	Expert 1	Test error rate = 0.057	Train error rate = 0.14 Test error rate = 0.14 (37 neurons)	Train error rate = 0 Test error rate = 0.14 (C=10, sigma=10)
Input A	Expert 2	Test error rate = 0.100	Train error rate = 0.26 Test error rate = 0.25 (25 neurons)	Train error rate = 0 Test error rate = 0.20 (C=10, sigma=10)
Input B	Expert 1	Test error rate = 0.057	Train error rate = 0.14 Test error rate = 0.14 (35 neurons)	Train error rate = 0 Test error rate = 0.14 (C=1, sigma=1000)
Input B	Expert 2	Test error rate = 0.100	Train error rate = 0.18 Test error rate = 0.25 (20 neurons)	Train error rate = 0 Test error rate = 0.21 (C=10, sigma=10)

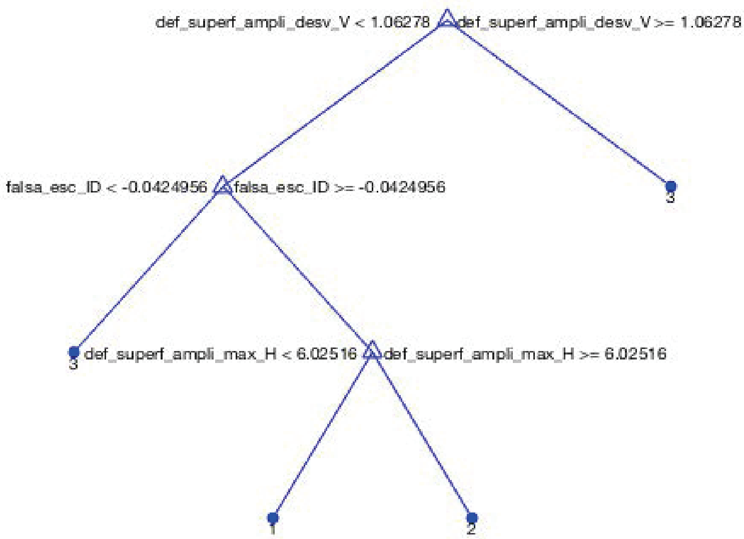

Fig. 1. CART model corresponding to the best prediction (error rate=0.057), simulating the criterion of Expert 1.

3 Conclusions

The best results correspond to the first model (CART), while neither MLP nor SVM were able to reproduce a proper classification criterion. This fact points out the linear separability of the classes of this classification problem. MLP and SVM are suitable for more complex problems, were non-linear relationships have to be established.
The results show a mathematical model can be used to achieve an objective automatic classification of slate slabs based on the information gathered by a computer vision system.

References

[1] Breiman, L., and Friedman, J. *Classification and Regression Trees.* Wadsworth, 1984.

[2] Martínez, J., Iglesias, C., Matías, J., Taboada, J., and Araújo, M. Solving the slate tile classification problem using a DAGSVM multiclassification algorithm based on SVM binary classifiers with a one-versus-all approach. *Applied Mathematics and Computation 230* (2014), 464–472.

[3] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review 65*, 6 (1958), 386–408.

[4] Vapnik, V. *Estimation of dependences based on empirical data.* Springer, 1982.