Materials and Design 82 (2015) 304–311

Contents lists available at ScienceDirect

Materials and Design

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

Prediction of mechanical strength of cork under compression using machine learning techniques

Ángela García [a], Ofélia Anjos [b,c], Carla Iglesias [a,*], Helena Pereira [c], Javier Martínez [d], Javier Taboada [a]

[a] Department of Natural Resources and Environmental Eng., University of Vigo, 36310 Vigo, Spain
[b] Instituto Politécnico de Castelo Branco, Escola Superior Agrária, Apartado 119, 6001-909 Castelo Branco, Portugal
[c] Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal
[d] Centro Universitario de la Defensa, Academia General Militar, 50090 Zaragoza, Spain

ARTICLE INFO

Article history:
Received 6 October 2014
Revised 17 March 2015
Accepted 21 March 2015
Available online 24 March 2015

Keywords:
Cork
Mechanical properties
Neural network
Multiple linear regression
CART
Cluster

ABSTRACT

In this study, the accuracy of mathematical techniques such as multiple linear regression, clustering, decision trees (CART) and neural networks was evaluated to predict Young's modulus, compressive stress at 30% strain and instantaneous recovery velocity of cork. Physical properties, namely test direction, density, porosity and pore number, as well as test direction were used as input. The better model was achieved when a classification problem was performed. Only compressive stress at 30% strain can be predicted with neural networks with an error rate of about 20%. The prediction of Young's modulus and instantaneous recovery velocity led to unacceptably high error rates due to the heterogeneity of the material.

1. Introduction

Cellular materials such as cork or foams have been used in many applications given their low density associated to a large compressibility, dimensional recovery, insulation properties, very low permeability to liquids and gases, and chemical stability and durability [1–3]. The exceptional properties of cork are due to its structure formed by small hexagonal prismatic closed cells that are stacked orderly base by base without intercellular voids [1].

Some authors have reported the mechanical behaviour of cork under static and dynamic load [2–12]. Some of these studies also concluded that the mechanical behaviour of cork is related to its physical properties, directly linked to the number and extension of defects.

The compression properties of cork are relevant for the performance of its application and vary mainly with density [2,3] and cell dimensions [1]. Cork presents a lower density, between 0.12 and 0.22 [3] due to its cellular structure, with cells with 85–90% of air inside [1].

Furthermore, the characteristics of the honeycomb structure of cork, crossed by pores of various sizes and where the cells have different dimensions, leads to highly variable mechanical properties, directly related to porosity [4,5,13]. This variability may be difficult to estimate because of the heterogeneous nature of the material: only for high quality cork was a good correlation between physical and mechanical properties reported [3].

The objective of this research work is to find a model to predict mechanical properties of cork under compressive stresses through a predictive model using previously measured physical properties. This model would provide researchers with information regarding the mechanical behaviour of cork without the need of performing destructive tests such as the usual compressive tests. To do so, the most popular methods of data mining and machine learning were used, from simple multiple linear regression to neural networks through clustering and decision trees (CART).

An artificial neural network (ANN) approach was successfully used to predict the compressive strength of lightweight and semi lightweight concretes with pumice aggregate subjected to high temperatures [14]. Tiryaki et al. work also with ANN and multiple linear regression model to predict optimum bonding strength of heat treated woods [15]. Regarding clustering analysis, it showed satisfactory results for the prediction of damage mechanisms in different composites [16,17].

Classification and regression trees have been applied to a wide variety of studies: for studying the surface in mining areas [18]; for modelling the behaviour of heavy metals in soil [19]; for studying corrosion defects in steels [20].

* Corresponding author.
E-mail address: carlaiglesias@uvigo.es (C. Iglesias).

http://dx.doi.org/10.1016/j.matdes.2015.03.038
0261-3069/

Regarding neural networks, several research works are found in the literature using this method in many fields: for the study of river status [21,22], for food analysis [23]; for the prediction of the mechanical properties of metal matrix composites [24,25], aluminium matrix composites [26] or polymeric composites [27]. However, only one recent study has been performed recently for the prediction of tensile properties of cork [28].

2. Materials and methods

2.1. Data

In order to construct the models using machine-learning techniques, 99 cork samples collected in a factory in Ponte de Sôr (Portugal) were used. The cork plates, of different grades were collected randomly after post-harvest water boiling operation and air-drying process that helped in the removal of growth stresses.

The test specimens were obtained from different cork planks with dimensions of 20 mm × 20 mm × 20 mm, dimensions that were chosen based on other studies [2,3,7] for comparisons of the results obtained.

The samples used in the study were already characterized in [2], and data of the mechanical and physical properties were selected from the group of data reported in [2]. According to that work, the test specimens were cut from different cork planks in cubes with 20 mm of edge oriented in the axial, tangential and radial directions of cork growth. The specimens were separated into three density classes: low density (0.13–0.15); mid density (0.15–0.19); and high density (0.19–0.25).

The porosity of specimens (given as a porosity coefficient, representing the area of pores in percentage over the total sample area) was evaluated by image analysis on all the specimen faces, i.e. tangential, transverse and radial sections, using a colour video camera and a Leica Qwin500 software with a 20× magnification. The number of pores was calculated as an average of the two tangential faces of the cork cubes.

The compression tests were carried out with applied loads in the axial, radial and tangential directions, at a constant crosshead speed of 2 mm min^{-1} (strain rate of 2×10^{-3} s^{-1}) up to a strain of 50%. Young's modulus was calculated from the average curve slope of the stress–strain curve between the stresses of 0.5 MPa and 0.9 MPa [2].

The physical and mechanical parameters selected for modelling in this study were: density (g/cm^3) (D); porosity (%) (P); number of pores (pore number/cm^2) (NP); Young's modulus (E); compressive stress at 30% strain (σ_{30}); instantaneous recovery velocity (mm/min) (IRV). All determination was made at cork moisture content of 9%.

2.2. Classical and machine learning techniques applied

Different mathematical techniques were used for the two approaches studied in this research work (initially, a regression problem; then, a classification problem): we started with a classical technique, multiple linear regression, and then tested machine learning techniques such as clustering, classification and regression trees (CART) and neural networks (MLP).

2.2.1. Multiple linear regression

Regression models describe the relationship between a dependent variable and one (or more) independent variable(s). The dependent variable is also called "response variable", while the independent variables are also called "explanatory variables". A multiple linear regression model relates a single response Y to a single predictor X for each observation. However in most problems there is more than one predictor, a fact that involves the use of multiple regression function.

A multiple linear regression model, where there are more than one explanatory variables, is defined as [29,30]:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where y_i is the i-th response, β_k is the k-th coefficient, β_0 is the constant term of the model, X_{ij} is the i-th observation of the j-th explanatory variable (with $j = 1, \ldots, p$) and ε_i is the noise in the i-th term. Also, the matrix of observations of the independent variables (X) is known as "design matrix".

In general terms, a multiple linear regression model can be expressed as:

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k f_k(X_{i1}, X_{i2}, \ldots, X_{ip}) + \varepsilon_i, \quad i = 1, \ldots, n \tag{2}$$

where $f(\cdot)$ is a scalar function of the explanatory variables X_{ij}. These functions $f(X)$ are of many types, including non-linear functions and polynomial functions. Linearity in this linear regression model refers to the linearity of the β_k coefficients. That is, the response variable y is a linear function of the β_k coefficients.

It is assumed that in the linear regression models the noise terms ε_i are unrelated, and their normal distributions are independent and identical, with zero mean and constant variance. It is also assumed that the responses y_i are unrelated.

Thus, the adjusted linear function is:

$$\hat{y}_i = \beta_0 + \sum_{k=1}^{K} \beta_k f_k(X_{i1}, X_{i2}, \ldots, X_{ip}), \quad i = 1, \ldots, n \tag{3}$$

where $\hat{y}_i$ is the estimated response. The coefficients are estimated so that the mean squared difference is minimized between the vector of prediction $\beta f(X)$ and the actual response vector (least squares method). With the above assumptions about the noise terms, these coefficients also maximize the similarity of the prediction vector. To establish an order of importance among the different variables of the problem the correlation coefficient is determined as follows:

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{S_Y^2} \tag{4}$$

where S_Y^2 is the variance of Y.

The strength of the linear relationship between the variables is quantified by the correlation coefficient R^2, which takes values from 0 to 1, with 1 being a perfect fit.

2.2.2. Clustering

This is a data mining technique based on the splitting of a group of objects into different subgroups (or clusters) mutually exclusive, so the similarity between the elements of a certain cluster is high and the similarity between elements from different clusters is low [31–35]. It can be applied to classification problems, and two main approaches are distinguished: hierarchical (the result is a nested series of partitions) and partitional (a single partition of the data is obtained).

A typical clustering involves the following steps [33]: (1) representation of the initial pattern, (2) definition of a pattern proximity measure according to the analyzed data, (3) initial clustering of the data, (4) assessment of results obtained. This process is usually repeated until the maximum similarity is achieved between the elements of the same cluster, and the minimum similarity is achieved between elements of different groups.