Food Research International 59 (2014) 1–7

Contents lists available at ScienceDirect

Food Research International

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

Geographical and genotypic segmentation of arabica coffee using self-organizing maps

Jade Varaschim Link [a,*], André Luis Guimarães Lemes [b], Izabele Marquetti [a], Maria Brígida dos Santos Scholz [c], Evandro Bona [a,b]

[a] Post-Graduation Program of Food Technology (PPGTA), Federal University of Technology — Paraná (UTFPR), P.O. Box 271, Via Rosalina Maria dos Santos — 1233, CEP 87301-899 Campo Mourão, PR, Brazil
[b] Food Department (DALIM), Graduation Program in Food Engineering, Federal University of Technology — Paraná (UTFPR), P.O. Box 271, Via Rosalina Maria dos Santos — 1233, CEP 87301-899 Campo Mourão, PR, Brazil
[c] IAPAR — Agronomic Institute of Paraná, Rodovia Celso Garcia Cid, Km 375, CEP 86047-902 Londrina, PR, Brazil

ARTICLE INFO

Article history:
Received 23 November 2013
Accepted 27 January 2014
Available online 4 February 2014

Keywords:
Green coffee
Unsupervised learning
Principal component analysis
Artificial neural networks

ABSTRACT

Several statistical methods have been developed in an attempt to reproduce the human capability of pattern recognition. Self-organizing maps (SOMs) are a type of artificial neural network (ANN) with unsupervised learning designed to examine the structure of multidimensional data. This study aimed to conduct a segmentation of the geographical and genotypic coffee grown in the coffee region of Paraná — Brazil using the SOM for cluster analysis. Fourteen arabica coffee genotypes from two different cities were collected (Paranavaí and Cornélio Procópio). Density, caffeine, chlorogenic acids, tannins, total and reducing sugars, proteins, and lipids of the green coffee beans were analyzed. Using these data, the SOM was able to discriminate the 14 genotypes and also segmentation of the geographical origin was observed. Reducing sugars, caffeine, and chlorogenic acid were the most important variables for separation of the region of cultivation of arabica coffee genotypes. It was concluded that the SOM was able to recognize the coffee genotypes and geographical origin using the chemical profile data.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Pattern recognition is a process by which a received signal is assigned to one class among a predetermined number of categories. Humans have excellence for learning and pattern recognition. Several statistical methods have been developed in an attempt to reproduce the human capability of pattern recognition (Bishop, 2006; Haykin, 2001).

Traditionally, for an exploratory study and data dimensionality reduction, principal component analysis (PCA) is one of the first multivariate methodologies chosen for cluster analysis. However, some disadvantages of the method are well known. First, it is assumed that the data can be described by linear combinations. Consequently, nonlinear systems will not be well represented. A second critical point in PCA is the quality of the result that can be influenced by discrepant samples. In addition, there is a possibility that after transformations, the number of significant components may still be high, which makes it difficult to extract useful information from the data (Borsato, Pina, Spacino, Scholz, & Filho, 2011; Melssen, Wehrens, & Buydens, 2006).

One methodology that can represent complex and nonlinear input–output relationships is the artificial neural networks (ANNs) (Bishop, 2006). ANNs are a set of techniques based on statistical principles, which are currently growing in food science to perform tasks of regression and pattern recognition. It is a methodology that employs massive interconnection of simple computing cells that has a natural propensity to store the experimental knowledge and make it available for use (Haykin, 2001; Lucia & Minim, 2006; Marini, 2009).

One kind of artificial neural network with unsupervised learning designed to examine the structure of multidimensional data is the self-organizing maps (SOMs) introduced by Kohonen (2001). In a SOM, the neurons are placed at the nodes of a lattice that is usually two-dimensional. The neurons become selectively tuned to various input patterns or classes of input patterns in the course of a competitive learning process, and thus a topographic map of the input patterns is formed in which the spatial locations of the neurons in the lattice are indicative of intrinsic statistical features contained in the input patterns. Because a self-organizing map is inherently nonlinear, it may thus be viewed as a nonlinear generalization of PCA (Haykin, 2001). Self-organizing maps have been largely applied in food science and technology in the characterization of strawberry varieties (De Boishebert, Giraudel, & Montury, 2006); identification of binary mixtures of Italian olive oils (Marini, Magrì, Bucci, & Magrì, 2007); characterization of rosemary samples according to their geographical origin (Tigrine-Kordjani, Chemat, Meklati, Tuduri, GIraudel & Montury, 2007); and aromatic pattern recognition of soluble coffee (Bona, Silva, Borsato, & Bassoli, 2012).

* Corresponding author. Tel.: +55 46 99156494.
E-mail address: jadejvl@hotmail.com (J.V. Link).

0963-9969/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodres.2014.01.063

The importance of coffee in the world economy is unquestionable, as it is one of the most valuable commodities traded in the world. Its cultivation, processing, transportation, and marketing provide millions of jobs around the world (SINDICAFÉ, 2012). Brazilian consumption of coffee increases every year, and between 2011 and 2012 consumption was 19.975 million bags, representing an increase of 3.05% over the previous period (ABIC, 2012). Besides the internal market, exports of green coffee from Brazil amounted to 116.63 thousand tons in 2012 (CECAFÉ, 2012).

There are two main species of coffee, *Coffea arabica*, also known as arabica coffee and *Coffea canephora* or robusta coffee (Farah, 2009; Higdon & Frei, 2006; Rubayiza & Meurens, 2005). These species present a very different chemical composition and the arabica coffee provides a better coffee beverage quality and flavor than robusta coffee (Farah, 2009). Among cultivars of arabica coffee has been found different levels of quality beverage due to genetic factors and environmental conditions in the place of cultivation (Bertrand, Boulanger, Dussert, Ribeyre, Berthiot, Descroix, et al., 2012; Jöet et al., 2010). The climatic conditions of coffee cultivation together with the genetic characteristics of the cultivars give special attributes to the beverage and could increase its value. However, it is essential to prove the geographical and genotypic origin of the cultivar using reliable methods (Borsato et al., 2011). Due to this fact, this study aimed to conduct a segmentation of the geographical and genotypic arabica coffee grown in the coffee region of Paraná — Brazil. For this purpose, a cluster analysis by SOM was performed utilizing the grain density and chemical data of coffee cultivars with wide genetic diversity grown in two contrasting climatic conditions.

2. Materials and methods

2.1. Samples of coffee

All coffee samples are genotypes of the species *C. arabica*. Samples of modern genotypes with wide variety of genetic background (IAPAR 59, IPR 97, IPR 99, IPR 100, IPR 101, IPR 102, IPR 105, Iapar 59 enx., IPR 105 enx. and Tupi) were collected in 2010 season. Traditional cultivars (Catuaí, Bourbon and Mundo Novo) were also collected at the same locations and same season. About 3 kg of cherry coffee of 14 genotypes (27 samples analyzed in duplicate) of coffee were collected at two locations in the coffee region of Paraná — Brazil: Paranavaí (23° 04′22″S 52° 27′ 54″ W; altitude 470 m, mean annual temperature 22–23 °C) and Cornélio Procópio (23° 10′51″S 50° 38′ 49″ W; altitude 658 m, annual average temperature 20–21 °C) and were transported to Agronomic Institute of Paraná (IAPAR) in Londrina—Paraná. The samples were immediately placed into wooden boxes with a mesh bottom and moved eight times per day until beans reached 11–12% moisture and then the samples were processed (removal of hull and parchment). For the chemical analysis, green coffee beans were frozen with liquid nitrogen and were ground in a mill disk (Perten 3600 model) with 0.6-mm final particle size (Brasil, 2011). The processed samples of coffee genotypes were subsequently used for analysis. Some samples were cultivated in different cities. The analyses to determine the chemical composition of the green coffee samples were performed at Plant Physiology Laboratory in IAPAR.

2.2. Density and chemical analysis

The density of the green bean was determined by the free-fall method, according to Buenaventura-Serrano and Castaño-Castrillón (2002). The results are expressed in $g \cdot mL^{-1}$. Caffeine was extracted with magnesium oxide and determined by the spectrophotometric method (IAL, 2008). Total chlorogenic acids were evaluated according to the methodology proposed by Clifford and Wight (1976). Total tannins were determined with Folin–Ciocalteu reagent using gallic acid as a standard (AOAC, 1990). Total sugar sucrose and reducing sugars were extracted with water at 70–80 °C and determined with the Somogyi and Nelson reagent (Southgate, 1976). Proteins and total lipids were determined by the respective methods proposed by the AOAC (1990).

2.3. Self-organizing maps

In this work, a two-dimensional SOM algorithm applied was introduced (Haykin, 2001). The principal goal of the SOM is to transform an incoming signal pattern of arbitrary dimension into a two-dimensional discrete map, and to perform this transformation adaptively in a topologically ordered fashion (De Boishebert et al., 2006; Haykin, 2001). In the Kohonen model (Fig. 1), each neuron in the lattice is fully connected to all the source nodes in the input layer. The algorithm responsible for the formation of the self-organizing map proceeds first by initializing the synaptic weights in the network. This can be done by assigning them small values (between 0 and 1) picked from a random number generator, thus, no prior order is imposed on the feature map. Once the network has been properly initialized, there are three

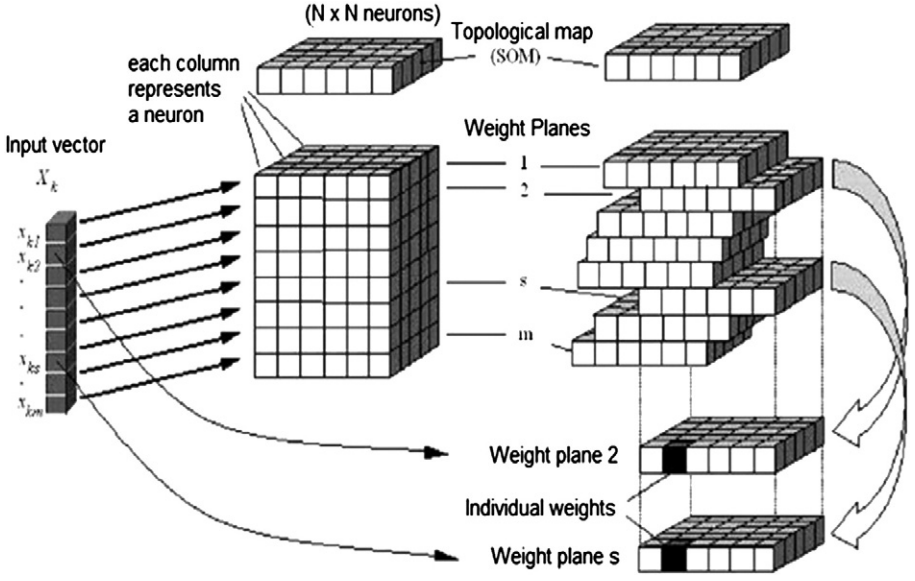

Fig. 1. Self-organizing map according to the model of Kohonen. *k* represents the number of input patterns, *m* is the number of input variables, and *N* is the number of neurons in each dimension (Marini, Zupan, & Magrì, 2005).

Download English Version:

https://daneshyari.com/en/article/6396954

Download Persian Version:

https://daneshyari.com/article/6396954

Daneshyari.com