Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Food Engineering 75 (2006) 1–10
JOURNAL OF FOOD ENGINEERING
www.elsevier.com/locate/jfoodeng

Feature extraction and classification of Chilean wines

N.H. Beltrán [a], M.A. Duarte-Mermoud [a,*], M.A. Bustos [a], S.A. Salah [a], E.A. Loyola [b], A.I. Peña-Neira [b], J.W. Jalocha [b]

[a] Department of Electrical Engineering, University of Chile, Av. Tupper 2007, Casilla 412-3, Santiago, Chile
[b] Department of Enology and Agroindustry, University of Chile, Av. Santa Rosa 11315, Santiago, Chile

Received 17 November 2004; accepted 14 March 2005
Available online 12 May 2005

Abstract

In this work, results of Chilean wine classification by means of feature extraction and Bayesian and neural network classification are presented. The classification is made based on the information contained in phenolic compound chromatograms obtained from an HPLC-DAD. The objective of this study is to classify different Cabernet Sauvignon, Merlot and Carménère samples from different years, valleys and vineyards of Chile. Different feature extraction techniques including the discrete Fourier transform, the Wavelet transform, the class profiles and the Fisher transformation are analyzed together with several classification methods such as quadratic discriminant analysis, linear discriminant analysis, K-nearest neighbors and probabilistic neural networks. In order to compare the results, cross validation and re-sampling techniques were used.

Keywords: Wine classification; Pattern recognition; Statistical classification; Bayesian classification; Wavelet transform; Fisher transform; Probabilistic neural networks; K-nearest neighbors

1. Introduction

During the last years, the Chilean wine industry has experienced a sustained growth, becoming one of the most important industries in the Chilean economy, reaching exportations of US$ 570 millions on 2000, US$ 590 on 2001 and US$ 610 on 2002. This growth is due to the incorporation of technology in this industry to compete in the international market.

In this work, we present the results of Chilean red wine classification, considering the varieties Cabernet Sauvignon, Merlot and Carménère from different valleys, years and vineyards. The classification is based on the information contained in phenolic compound chromatograms obtained from an HPLC-DAD.

In most of previous wine classification papers the concentration of specific compounds are the main variables on which the classification is based. Typically this concentrations are obtained from liquid and gas chromatography or other techniques, and correspond to characteristics such as major acids (Cabezudo, Herraiz, & De Gorostiza, 1983; Etievant, Schlich, Cantagrel, Bertrand, & Bouvier, 1989), anthocyanins (Aires de Sousa, 1996; Berente, García, Reichenbacher, & Danzer, 2000), free amino acids (Vasconcelos & Chaves das Neves, 1989), biogenics amines (Csomos, Heberger, & Simon-Sarkadi, 2002), isotropic ratios (Kosir, Kocjancic, Ogrinc, & Kidric, 2001), aromas (Weber et al., 1999; García, Reichenbacher, & Danzer, 1998), phenolic composition (García-Parrilla, González, Heredia, & Troncoso, 1997), color (Almela, Javaloy, Fernández-Lopez, & López-Roca, 1996; Ortiz-Fernández, Herrero-Gutiérrez, Sánchez-Pastor, Sarabia, & Iñiguez-Crespo, 1995) etc. Classification has then been done using directly the information provided by the sensors using a wide

* Corresponding author. Tel.: +56 2 678 4213; fax: +56 2 672 0162.
E-mail address: mduartem@cec.uchile.cl (M.A. Duarte-Mermoud).

0260-8774/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2005.03.045

variety of methods ranging from statistical (Csomos et al., 2002; Kallithraka et al., 2001; Kaufmann, 1997; Tzouros & Arvanitoyannis, 2001) to artificial neural networks and genetic algorithms (Aires de Sousa, 1996; Almela et al., 1996; Beltrán et al., 2005; Ortiz-Fernández et al., 1995; Sun, Danzer, & Thiel, 1997).

In a chromatogram the compound concentration depends on the area under the peak that appears at the time when the column liberates it, known as retention time. This time depends mainly on factors like the temperature gradient applied to the sample, column aging and compound type. The commonly used methodology in order to identify specific compounds is to establish experimental conditions and then associate a specific compound to a retention time using standard chromatographic patterns. This approach, besides of requiring a previous identification of the analyzed compounds, needs to identify which of them are the most important for a specific wine characterization, which is an open problem.

In this work, a different approach is presented, which does not requires previous compound identification, because the classification is made using the whole information contained in the chromatogram signal, instead of only the areas of some interesting peaks. The difficulty of this approach is that commonly, the chromatographic information is characterized for having a huge data volume, making a direct approach with classification techniques, like Discriminant Analysis or Neural Networks classifiers, almost impossible, because of the denominated course of dimensionality (Fukunaga & Hayes, 1989).

Nevertheless using signal analysis tools and feature extraction techniques before the classification task, it is possible to reduce the dimension of the data and to obtain wine classification rates of about 95%.

2. Description of the experimental data

In this study 172 Chilean red wine liquid chromatograms were analyzed. Chromatograms were obtained from confident samples of 80 Cabernet Sauvignon, 35 Merlot and 57 Carménère wines, cultivated in Maipo, Rapel, Curicó, Maule and Itata valleys in the central zone of Chile, between the years 2000 and 2001.

The information contained in the chromatograms, corresponds to phenolic compounds of small molecular weight obtained through a high performance liquid chromatogram (HPLC) attached with an aligned photodiode detector (DAD) (Peña-Neira, Hernández, García-Vallejo, Estrella, & Suarez, 2000). The equipment used in this study is a liquid chromatograph Merk-Hitachi model L-4200 UV–Vis Detector with internal pump, and a thermostat column holder. The column used is a Novapack C18, of 300 mm length and 3.9 mm of internal diameter. To separate different phenolic compounds the following solutions were used as solvents:

(A) 98% H_2O, 2% acetic acid,
(B) 78% H_2O, 20% acetonitrile 2% acetic acid.

The gradient used in this tests was, 0–55 min, 100% A at 1 ml/min; 55–57 min, 20% A and 80% B at 1 ml/min; 57–90 min, 10% A and 90% B at 1.2 ml/min.

Each digitalized chromatogram has a length of 6751 points and some peaks can be identified as an specific phenolic compound. These compounds have been widely studied and identified by chemical investigators and agronomic researchers (Alamo, 2002; Muñoz, 2002; Peña-Neira et al., 2000). Figs. 1–3 show typical profiles of the Chilean Merlot, Cabernet Sauvignon and Carménère red wines obtained from an HPLC-DAD suitably normalized.

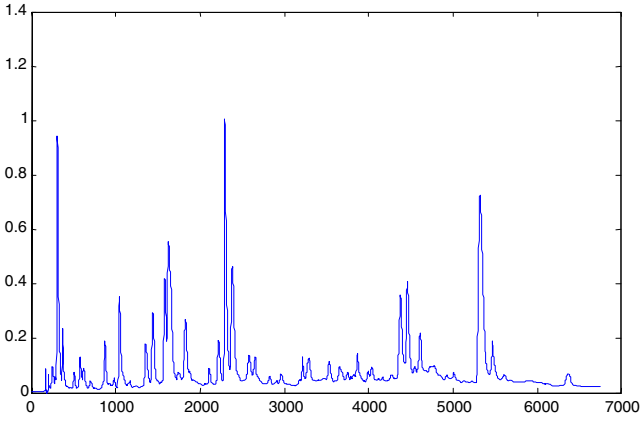

Fig. 1. Typical Chilean Merlot red wine phenolic normalized chromatogram.

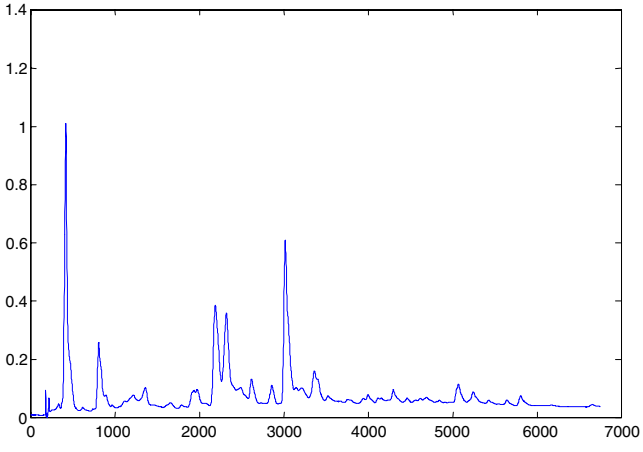

Fig. 2. Typical Chilean Cabernet Sauvignon red wine phenolic normalized chromatogram.