Computers and Electronics in Agriculture 117 (2015) 186–193

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Investigating the use of gradient boosting machine, random forest and their ensemble to predict skin flavonoid content from berry physical–mechanical characteristics in wine grapes

Luca Brillante [a,*], Federica Gaiotti [a], Lorenzo Lovat [a], Simone Vincenzi [b], Simone Giacosa [c], Fabrizio Torchio [c], Susana Río Segade [c], Luca Rolle [c], Diego Tomasi [a]

[a] CRA-VIT Council for Agricultural Research and Economics, Viticulture Research Center, Conegliano, TV, Italy
[b] University of Padova, Centro Interdipartimentale per la Ricerca in Viticoltura ed Enologia, Legnaro, PD, Italy
[c] University of Turin, Dipartimento di Scienze Agrarie, Forestali e Alimentari, Grugliasco, TO, Italy

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form 1 July 2015
Accepted 5 July 2015
Available online 24 August 2015

Keywords:
Random forest
Gradient Boosting Machine (GBM)
Wine-grape
Flavonoids
Texture analysis

ABSTRACT

Flavonoids are a class of bioactive compounds largely represented in grapevine and wine. They also affect the sensory quality of fruits and vegetables, and derived products. Methods available for flavonoid measurement are time-consuming, thus a rapid and cost-effective determination of these compounds is an important research objective. This work tests if applying machine learning techniques to texture analysis data allows to reach good performances for flavonoid estimation in grape berries.

Whole berry and skin texture analysis was applied to berries from 22 red wine grape cultivars and linked to the total flavonoid content. Three machine-learning techniques (regression tree, random forest and gradient boosting machine) were then applied. Models reached a high accuracy both in the external and internal validation. The R^2 ranged from 0.75 to 0.85 for the external validation and from 0.65 to 0.75 for the internal validation, while RMSE (Root Mean Square Error) went from 0.95 mg g^{-1} to 0.7 mg g^{-1} in the external validation and from 1.3 mg g^{-1} to 1.1 mg g^{-1} in the internal validation.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Flavonoids are a group of secondary metabolites widely distributed in plants, which greatly affect the sensory and nutritional quality of fruits and vegetables (Harnly et al., 2006). They represent a huge portion of soluble phenols present in grapevine (Braidot et al., 2008). Flavonoids are among the most important compounds for the quality of red wine grapes because of their effect on wine sensory attributes (Ristic et al., 2010 is an example) and aging. The concentration of these compounds in wine depends, among other factors, on the quantity originally present in grapes (González-Neves et al., 2004). In the last ten years, flavonoids have received a very great attention from both researchers and the general audience because of their beneficial effect on human health (Yao et al., 2004). They have shown antioxidant (Lourenço et al., 2008), hypocholesterolemic (Gonzalez et al., 2015) and anti-inflammatory effects (Noll et al., 2009). Their nutraceutical properties are exploited in fresh table grapes, in pharmaceutical and cosmetic products derived from grape, and are a very appealing argument for wine marketing purposes.

Red grapes are richer in flavonoids than white ones, but their biosynthesis and concentration greatly depend on cultivar, vineyard practices, soil and climate (Koundouras et al., 2006). Grape maturity, and therefore the harvest date, is also another very important parameter because quantitative and qualitative modifications of tannins and anthocyanins (the two most represented flavonoid families in grape) happen during ripening (Kuhn et al., 2013).

Different methods based on spectrophotometry, chromatography, and mass spectrometry are usually used for the determination of flavonoids in fruits and vegetables (see Ignat et al., 2011 for a generic review and Lorrain et al., 2013 for the case of grapes and wine). Regarding grape analysis, these methods are all very accurate but they often require sample preparation and long analysis times. The problem is especially the time required for the extract preparation and purification, which has to be made by hand and can require berry peeling, solvent extractions, and

Abbreviations: BW, Berry Weight; CART, Classification And Regression Tree; Co, cohesiveness; Ch, chewiness; E_{sk}, skin young's modulus; F_{sk}, skin break force; *G*, gumminess; GBM, gradient boosting machine; *H*, hardness; *R*, resilience; RF, Random Forest; RFE, Recursive Feature Elimination; RT, Regression Tree; *S*, springiness; Sp_{sk}, skin thickness; SW, skin weight; W_{sk}, skin break energy; TA, Texture Analysis; TF, total flavonoids index.

* Corresponding author.
E-mail address: brillanteluca@live.it (L. Brillante).

http://dx.doi.org/10.1016/j.compag.2015.07.017
0168-1699/© 2015 Elsevier B.V. All rights reserved.

other manipulations that strongly increase costs and limit the number of acquirable data. Industry and research will greatly benefit from a rapid and cost effective method to obtain a faster screening of flavonoids in grapes. Such a method is at today lacking, although recently great advances have been made in this field by the use of Near InfraRed (NIR) spectroscopy coupled to chemometrics, in particular using partial least squares (PLS) regression models (Ferrer-Gallego et al., 2011; Rolle et al., 2012a; Cozzolino 2015).

During grape ripening, berries change not only their chemical composition, but also their mechanical properties: they soften, become less resilient, and the skin generally harden (Rolle et al., 2012b). In industry, these textural modifications are currently evaluated by sensory panels to help in the choice of the harvest date. Texture Analysis (TA) has shown to be an effective instrumental technique for an accurate evaluation of physical–mechanical characteristics of grapes (Letaief et al., 2008; Giordano et al., 2013; Battista et al., 2015). It is cost-effective as it does not require long times and reagents for sample preparation and analysis.

Although flavonoids and texture parameters belong to different grape properties, their values are both influenced by the berry ripening process. The phenolic ripeness of grape skin was found to be well assessed when the TA values were used (Río Segade et al., 2008), but the possibility of a predictive model has been never investigated, and neither an evaluation of possible chemometrics approaches to these parameters exists. A model linking the differences in berry mechanical properties and chemical composition induced by the grape ripeness could be an alternative to NIR methods for rapidly assessing the flavonoid contents at the berry level.

TA data are different from those obtained with NIR. In the first method, the number of measured parameters available as predictors is limited, and it is generally lower than the number of observations, i. e. the dataset is in a long format. Conversely, NIR datasets are wider, the number of wavelengths available as predictors is large and therefore PLS, a regression algorithm well suited to these situations, has been extensively applied (Cozzolino, 2015). With the reduced number of predictors present in TA, other learning algorithms could be effectively applied as an effort to better exploit the available information.

In this work, we will evaluate the use of regression trees and of two ways of combining them in order to achieve greater performances in predictions: Random Forest, RF (Breiman, 2001), and gradient boosting machine, GBM (Friedman, 2001). RF has shown to be a state-of-the art method, allowing the highest accuracy, but it is still not widespread to date. According to a recent review by Scott et al., 2013 for chemometric classification problems (286 reviewed papers), RF is used in only 4.5% of the articles where machine-learning algorithms are applied. The same source evidences that boosting algorithm is even less used (1%).

The aim of the work was to evaluate different chemometric approaches in the evaluation of data obtained from parameters influenced by the grape ripening process, such as berry mechanical properties data and flavonoid content in berry skins. For this, the performances of RF and GBM algorithms were compared on a large dataset composed of approx. 800 berries belonging to 22 grapevine cultivars, their suitability for flavonoid content prediction in grape berries was evaluated on the basis of mechanical properties, and an informal explanation of the underlying algorithms was suggested. Furthermore, a predictive model was also developed. This approach could be used as an example for other compounds and fruits.

2. Materials and methods

2.1. Grape sampling

Grapes from 22 red grapevine cultivars (*Vitis vinifera* L.) were sampled in the CRA-VIT experimental collection (1.2 ha) located in Susegana (TV), Veneto Region (North-East Italy), in 2010 and 2011. Vines were 15 years old, grafted on SO4 rootstock (interspecific cross between *Vitis riparia* Michx. and *Vitis berlandieri* Planch.), and planted at 3.0 m between rows and 1.5 m between vines. They were Sylvoz pruned and trained with a vertical shoot position system. For each cultivar, samples were composed of approx. 3 kg of grape berries, which were picked up randomly from ten vines. In order to successfully compare berries at ripeness with adequate sugar content, the berries were calibrated using a densimetric method by berry flotation in different saline solutions (Rolle et al., 2011). This study was carried out only on the berries with sugar contents comprised between $183 \pm 8\ g\ L^{-1}$ and $217 \pm 8\ g\ L^{-1}$ corresponding to $11.0 \pm 0.5\%$ (v/v) and $13.0 \pm 0.5\%$ (v/v) potential alcohol, respectively.

The sorted berries were visually inspected before analysis; those with damaged skins were discarded. For each variety studied, a sub-sample of 36 sorted berries (therefore a total of 792 berries for all cultivars together) was randomly selected for the determination of the physical–mechanical properties and then for the flavonoid content. As described in the successive section, single berries measurements were then averaged by three to compose a single sample for predictive modeling.

2.2. Physical and mechanical properties

Grape berries were singularly weighed, with an analytical laboratory balance Radwag AS 220/X (Radwag, Radom, Poland), and then a Texture Profile Analysis (TPA) non destructive mechanical test was performed for each of them as described by Letaief et al., 2008. It allowed the measurement of berry hardness (N, as *H*), cohesiveness (adimensional, as Co), gumminess (N, as *G*), springiness (mm, as *S*), chewiness (mJ, as Ch) and resilience (adimensional, as *R*). A puncture test (Letaief et al., 2008) was then carried out on the same berries taken singularly to measure skin break force (N, as F_{sk}), skin break energy (mJ, as W_{sk}) and skin resistance to axial deformation ($N\ mm^{-1}$, as E_{sk}). All these measurements were performed on the equatorial position of whole berry, while skin thickness (μm, as Sp_{sk}) was measured in the skin after manual removal from the pulp with a razor blade (Letaief et al., 2008; Río Segade et al., 2011a). Analyses were made with a Universal Testing Machine (UTM) TAxT2i texture analyzer (SMS-Stable Micro Systems, Godalming, Surrey, UK) equipped with a 5 kg load cell and a HDP/90 platform. A SMS P/35 flat probe under 25% deformation, with a waiting period of 2s between the two compressions and a speed of $1\ mm\ s^{-1}$, was used for the TPA test. A SMS P/2N needle probe, with a test speed of $1\ mm\ s^{-1}$ and a penetration depth of 3 mm, was used for the puncture test. A SMS P/2 flat probe, with a test speed of $0.2\ mm\ s^{-1}$ was used to measure Sp_{sk}. All data were acquired at 400 Hz and evaluated using the Texture Expert Exceed software, version 2.54.

2.3. Skin flavonoid content

After the skin thickness test, each berry skin was individually immersed for 4 h in 5 mL of a buffer solution containing 12% v/v ethanol, $2\ g\ L^{-1}$ of $Na_2S_2O_5$, $5\ g\ L^{-1}$ of tartaric acid and adjusted to pH 3.20 with NaOH (Di Stefano and Cravero, 1991). Each skin was then homogenized at 8000 rpm for 1 min with an Ultraturrax T18 (IKA Labortechnik, Staufen, Germany), and the extract was centrifuged for 10 min at $3500 \times g$ and 20 °C. The supernatant was then used for analysis after dilution with an ethanolic solution of HCl (70:30:1, ethanol:water:HCl, v/v) (Di Stefano and Cravero, 1991). Total flavonoid index (TF) was determined by a spectrophotometric method, reading the absorbance at 280 nm, using an Uvmini-1240 PC spectrophotometer (Shimadzu Scientific

Download English Version:

https://daneshyari.com/en/article/6540655

Download Persian Version:

https://daneshyari.com/article/6540655

Daneshyari.com