IL NUOVO CIMENTO **47 C** (2024) 298
DOI 10.1393/ncc/i2024-24298-1

COMMUNICATIONS: SIF Congress 2023

Fast and efficient feature importance calculations for EVOO spectroscopic data

L. ORTENZI([1]), S. VIONLINO([2]), S. FIGORILLI([2]), F. PALLOTTINO([2]), F. ANTONUCCI([2]), L. MOSCOVINI([2]) and C. COSTA([2])

([1]) *Dipartimento di Scienze Agrarie e Forestali, Università degli Studi della Tuscia - Via San Camillo de Lessis s.n.c., 01100 and Via Angelo Maria Ricci 35a 02100 Rieti, Italy*
([2]) *Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria (CREA), Centro di Ricerca Ingegneria e Trasformazioni Agroalimentari - Via della Pascolare 16, 00015 Monterotondo (Rome), Italy*

received 22 February 2024

Summary. — Chemometrics via artificial neural networks is nowadays the state of the art for food analysis. Here we consider an application aiming to distinguish European (EU) extra vergin olive oil (EVOO) from non EU one. Though the interpretability of neural network is still an open issue, here we propose a simple and transparent method to "open the box" and showing the relative importance of the spectral wavelengths in determining the classification output.

1. – Introduction

Traceability of extra virgin olive oil (EVOO) aims to guarantee the authenticity of the product and protect the consumer within a system that is able to certify the traceability information [1]. There are several techniques for verifying the traceability of oils, but often some of these methods have high costs and need suitable personnel to analyze and interpret the data [2]. In contrast, there are techniques such as VIS-NIR spectroscopy, which overcome these problems associated with traditional methods but allow for analysis of food authenticity and verification of traceability, being very low-cost and ready-to-use [1]. In the work reported by Al Riza [3] VIS-NIR spectroscopy was used to identify EVOO cultivars and geographical origin, demonstrating that at least 3 excitation wavelengths are required to obtain 5 variables capable of explaining more than 95% of the sample variance with 4 main components. As a result, multiple excitation fluorescence techniques can be used for authentication of EVOO cultivars and geographic origins. A non-destructive NIR spectroscopy method was also used in ref. [4] to assess the quality of EVOO. In particular, two portable low-cost NIR instruments were used that allowed the detection of atypical olive oils and obtained excellent precision, in particular for the predictions of palmitic and oleic acid. The discrimination between different classes of EVOOs is usually performed by discriminating their VIS-NIR relative absorbance through an artificial neural network [5]. Indeed, given the presence of overtones and several optical source of noise VIS-NIR spectroscopy is based on the recognition

of what is "called the spectral signature" [6]. Though the interpretability of artificial neural networks is still an open issue, a step forward the actual discrimination criterion can be made based on what is called "variable impact" or "feature importance" [7]. One way to calculate this quantity consists of evaluating the sensitivity of input variables to controlled noise [8]. Albeit effective, this procedure implies iteration on all the samples and on all the variables involved.

In this work we compare the feature important obtained through the well-established procedure described in ref. [8] with a more efficient procedure based on the analysis of the weight of the artificial neural network. For the calculations we consider a dataset of 203 EVOO samples classified as "European" and "Non European" for which the shallow neural network of 50 nodes in the hidden layer showed no missclassification both on the training and on the test set [5].

2. – Shallow neural network architecture

A shallow neural network is an artificial neural network made of one input layer, an output layer and only one hidden layer [9]. The input layer is constructed in the following way. The input variables associated with m samples of olive oil, whose spectrum is represented by n intensities corresponding to n sampled wavelengths are arranged row-wise into a matrix $\hat{X}_{m,n}$.

First the input matrix $\hat{X}_{m,n}$ is multiplied by the weights of the input layer matrix $\hat{W}^{(1)}_{n,h}$ with n rows and h columns, than a bias matrix a bias matrix $\hat{B}_{m,h}$ is added to $\hat{W}^{(1)}_{n,h}$ and an activation function $g^{(1)}$ is applied to the result. The input layer $z^{(1)}$ is then described by the following equation:

$$z^{(1)}_{m,h} = g^{(1)}(\hat{X}_{m,n}\hat{W}^{(1)}_{n,h} + \hat{B}_{m,h}). \tag{1}$$

The bias matrix has actually only h parameters since is obtained by stacking a bias vector $b_{1,h}$ row-wise m times as shown in fig. 1.

Similarly, the hidden layer, is described by the following equation:

$$z^{(2)}_{m,o} = g^{(1)}(z^{(1)}_{m,h}\hat{W}^{(2)}_{h,o} + \hat{C}_{m,o}). \tag{2}$$

where the bias matrix $\hat{C}_{m,o}$ is obtained in a similar way as $\hat{B}_{m,h}$. The above formalism expresses the fact the activation functions $g^{(1)}$ and $g^{(2)}$ are applied to all the elements

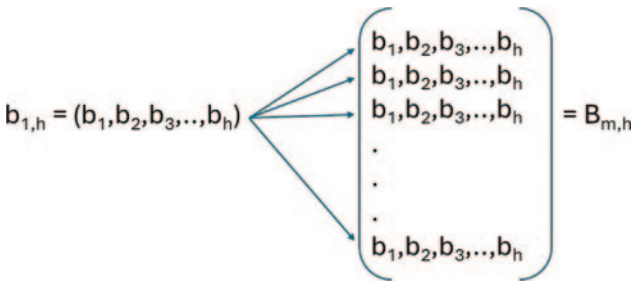

Fig. 1. – The bias matrix has only h parameters and is obtained by stacking a bias vector $b_{1,h}$ row-wise m times.

of the matrices in their argument. The matrix elements of $\hat{W}^{(1)}$, $\hat{W}^{(2)}$, $\hat{B}$ and $\hat{C}$ are obtained by minimizing an appropriate cost function [9] through a procedure known as back propagation algorithm [10] which adapts the output of eq. (2) to the vector $\hat{Y}$ made of the known classes of the EVOO samples.

3. – Feature importance and analysis of $\hat{W}^{(1)}$

Once converged the neural network is uniquely defined by the matrix elements of $\hat{W}^{(1)}$, $\hat{W}^{(2)}$, $\hat{B}$ and $\hat{C}$ and to each EVOO samples, characterized by its corresponding optical spectrum $x^{(i)} = (x_1^{(i)}, x_1^{(i)}, \ldots, x_n^{(i)})$, the net assigns a number $p^{(i)}$ representing the predicted class. The feature importance FI of the j-th variable represents the sensitivity of the model to a change in the j-th frequency ω_j of the spectrum. $FI(\omega_j)$ can be estimated by calculating the variation of the output related to the variation of the intensity of the j-th frequency ω_j accross the m sample s [11]. This procedure however, implies iteration on the n frequencies and on the m samples with a consequent important computational load. Here we propose a different approach consisting in calculating the absolute value of the matrix element of $\hat{W}^{(1)}$ and calculating its maximum over the columns. This procedure is based on the fact that the matrix $\hat{W}^{(1)}$ works as a gate for the input variables. If its element are large enough they allow the signal to propagate through the net and eventually giving rise to the output $p^{(i)}$, otherwise it gets stoped at the entrance. Figure 1 shows a comparison between FI calculated according to refs. [8] and [11] and the one shot variable impact defined as

$$FI^{\text{one shot}}(j) = \max_h(|\hat{W}^{(1)}_{j,h}|), \tag{3}$$

where the absolute value must be considered as applied element wise. Both quantities are normalized to the maximum. We tested this procedure on a dataset where a trained neural network run with 100% accuracy [5]. In this way we are sure to have the maximum signal to noise ratio since the model parameters are perfectly adapted to the expected output.

The two quantities reported in fig. 2 correlate very well with a Pearson coefficient $R^2 = 0.79$ with the following regression equation:

$$FI^{\text{one shot}} = 1.06\ FI + 0.11. \tag{4}$$

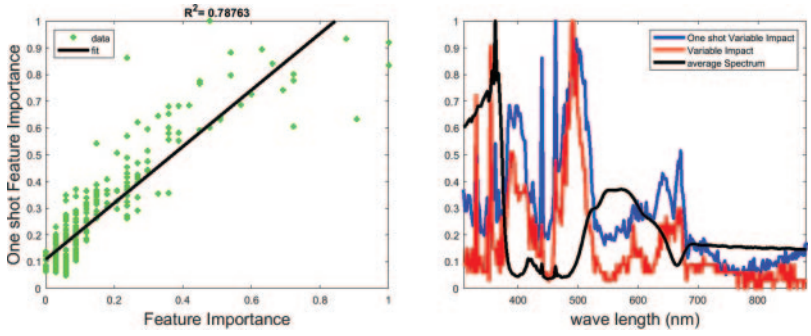

Fig. 2. – One shot feature importance *vs.* feature importance calculated as in ref. [8].

They show peaks in the same positions corresponding to the absorbance peaks of poliphenols and carotenoids acually present in the EVOO [12,13]. The difference between FI and $FI^{\text{one shot}}$ resides in the height of the peaks which could be relied to a normalization issue. The plotted quantities are indeed both normalized to the maximum but and this could not be the most appropriate normalization. However, the investigation of this issue is beyond the scope of the present work.

4. – Conclusion

In the present short communication we have shown a comparison between the straightforward feature importance and the one shot feature importance calculated on a dataset of EVOO samples classified as "European" and "Non European". The classification accuracy, defined as the ratio between the number of correctly classified samples and the total number of samples, is 1 both on the training and on the test set. Albeit no formal theory exists yet linking the two quantities in fig. 2, the correlation among them and the construction procedure allows for the interpretation of the cassification results obtained at least when the accuracy, is close to one. Note that the difference in computational time is reduced by about 3000 times on a 32 GB ram. This allows to save a significant amount of computational time and resources when training the classification algorithm on a given EVOO dataset. Generally speaking the method reported here is of great importance when the number of variables considered is large.

* * *

LO acknowledges the framework of the Ministry of University and Research (MUR) initiative "Departments of Excellence" (Law 232/2016) DAFNE Project 2023-27 "Digital, Intelligent, Green and Sustainable (acronym: D.I.Ver.So.)".

REFERENCES

[1] Violino Simona *et al.*, *Foods*, **9** (2020) 624.
[2] McVey Claire *et al.*, *Trends Food Sci. Technol.*, **118** (2021) 777.
[3] Al Riza Dimas Firmanda *et al.*, *Food Control*, **121** (2021) 107604.
[4] Arroyo-Cerezo Alejandra *et al.*, *Food Chemistry*, **430** (2024) 137043.
[5] Ortenzi Luciano *et al.*, *Artificial intelligence approaches for fast and portable traceability assessment of EVOO*, in *2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)* (IEEE) 2023, pp. 1–5, https://doi.org/10.1109/COINS57856.2023.10189267.
[6] Ozturk Samet *et al.*, *J. Food Eng.*, **341** (2023) 111339.
[7] Violino Simona *et al.*, *Eur. Food Res. Technol.*, **248** (2022) 1011.
[8] *Palisade*, Version v1.11.9, https://palisade-crypto.org/.
[9] Goodfellow Ian, Bengio Yoshua and Courville Aaron, *Deep Learning* (MIT Press) 2016, http://www.deeplearningbook.org.
[10] Rumelhart David E., Hinton Geoffrey E. and Williams Ronald J., *Nature*, **323** (1986) 533.
[11] Navarro Alejandra *et al.*, *Plant Methods*, **18** (2022) 45.
[12] Essafi H. *et al.*, *Acta Aliment.*, **48** (2019) 384.
[13] Violino Simona *et al.*, *Eur. Food Res. Technol.*, **247** (2021) 1013.