Non agressive orange acid and sugar indexes estimation system

Tomás J. Arnau, José M. Sotoca, Filiberto Pla

Institute of New Imaging Technologies, Jaume I University, Castellón, Spain
Corresponding author: tarnau@uji.es

Abstract: *This work presents a non aggressive methodology for the measurement of acid and sugar indexes of oranges using spectral information. Our method tries to obtain the value of sugar and acid from a fruit, only analyzing the spectrum of light transmitted through the fruit. We apply support vector machine regression to estimate the parameters that permit us to assess the sugar and acid index content of fruits.*

1 Introduction

Nowadays, the worldwide market makes easier to find what you want and bring it to you. For this reason you need to be competitive to sell your products. In the fruits market you can be more competitive if you know beforehand the acid and sugar content of fruits, because in this case you can be more selective with the destination of the product, for instance, you can use the more sweet oranges to be consumed in the fresh market and the less sweet for orange juice.

It is possible to do a classification of the oranges attending to the level of sugar and/or acid for a quality control system, for example. The conventional analysis that have been used so far to know the acid and sugar indexes of oranges are highly invasive methods. It means that you have to do orange juice and analyze it in a chemistry laboratory (LI *et al.* 1996).

A new non aggressive methodology was developed for the measurement of acid and sugar indexes of the oranges using spectrophotometer (NICOLAÏ *et al.* 2007). This method tries to obtain the value of sugar and acid from an orange, only having the results of the spectrum of light that crosses it. There are some papers on nondestructive measurement fruit and vegetable quality with spectroscopy (STEINMETZ *et al*. 1995, TSUCHIKAWA *et al.* 2004). The aim of the present study is to report the analytical results of our method on validation by transmittance of oranges for the estimated ones.

We have tried to obtain a method with non parametric regression and without a model. Thus we have used the support vector machine (SVM) (HSU *et al.* 2003). This is a useful and popular technique for data classification and regression. A classification task usually involves with training and testing data which consist of some data instances.

ISSN 0947-7314
Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V. (ATB)

The goal of SVM is to produce a model that predicts the target value of data instances in the testing set, which are giving only the attributes. Applying this to our case, we have done measurements of the spectrum of light that crosses the orange. As a validation, we made orange juice in the laboratory with the oranges to obtain the values of the acid and sugar. We have applied the SVM to this relationship database.

2 Materials and methods

Samples

The samples of our system are oranges taken from trees located at the east of Spain, in fact, all of them from an area next to Castellón. Our set of oranges was composed by three hundred oranges harvested in subsets of ten oranges of several varieties picked up weekly during the ripening season.

System

The system is composed by an illuminating box with twelve halogen lights inside, each one with a consumption power of 150 watts; the measure system used is a Hamamatsu MiniSpectrometer TM-VIS/NIR CCD-ASeries with an achromatic collimated lens; the spectral response range of the spectrometer is rated between 320 and 1000 nanometers. The method consists of measuring the light signal that crosses the orange. To do this, we use the device described above trying to obtain only the light that crosses the orange. For this purpose, a torus made of foam was used. Over this torus the orange is put, and the extreme of the sensor is located below the fruit (**Figure 1**).

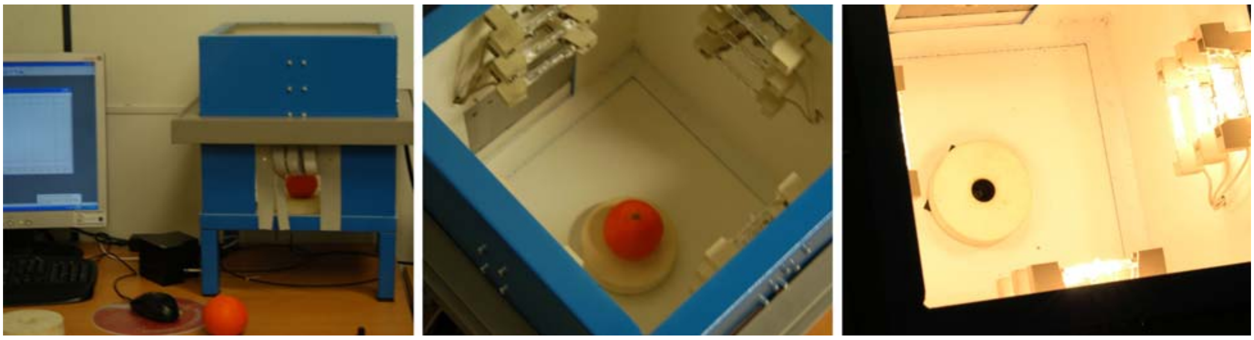

Figure 1: An overview of the system

After the signal was recorded, orange juice of the samples was obtained to measure the acid and sugar empirically. To do that, the standard technique was used to know the acid index of an orange. And, to know the sugar index we have used a refractometer.

Database Analysis

Given the spectrum signal and the corresponding recorded sugar and acid indexes, SVM regression was applied to estimate the parameters. It means that we have two data sets (sugar and acid) with 2048 values of wavelength and the values of sugar and acid indexes relatives to that signal for every orange. With these two databases we have done several tests trying to improve the results: first we have applied SVM to the data with half set of the data for training and the other half for the test. Next, we have done the same using a 10-fold cross validation data system, having 10% of data for training and 90% for test.

3 Results and discussion

To assess the reproducibility of our technique, we have analyzed about 300 oranges in several months, since the beginning of the period of ripeness to the end of it. We have built a database that associates the values of the spectrum and the data obtained in the analysis of the acid and sugar. The estimation of relationship of those values is realized by SVM regression.

To increase the accuracy of the results, we have tried several methods including Hold-Out and 10-Fold Cross Validation obtaining similar results in all cases. The absolute error of our method is 1.24 in a range between 5.61 for the minimum value and 17.14 for the maximum value of the acid; and for the sugar, the absolute error is 0.97 in a range between 11.0 for the minimum and 16.0 for the maximum.

Figure 2 shows the relationship between the obtained signal by the SVM and the original signal (red color). The blue signal is for the training set (fifty per cent of data), and the green one is for the test data.

In **Figure 3** appears the relationship of the observation data versus prediction data after SVM regression was applied. The optimum case would be if all the dots where located in a line that cuts the graph by two empty parts. Therefore, all the predicted data fits with its corresponding observed data.

ISSN 0947-7314
Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V. (ATB)

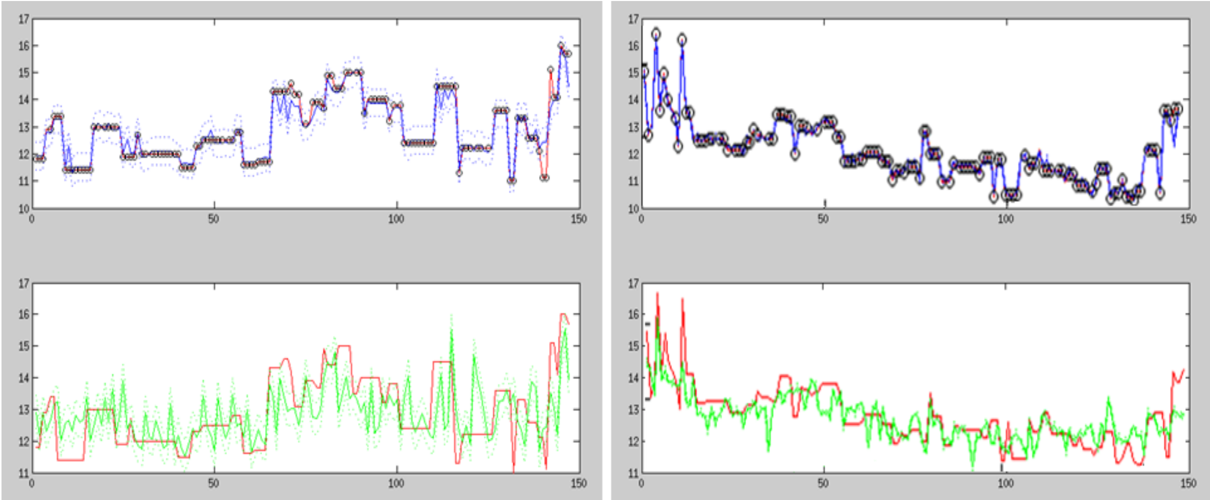

Figure 2: In these pictures the red line represents the original signal. Left graphs represent training (up) and test (down) signals for sugar indexes. Right are for acid.

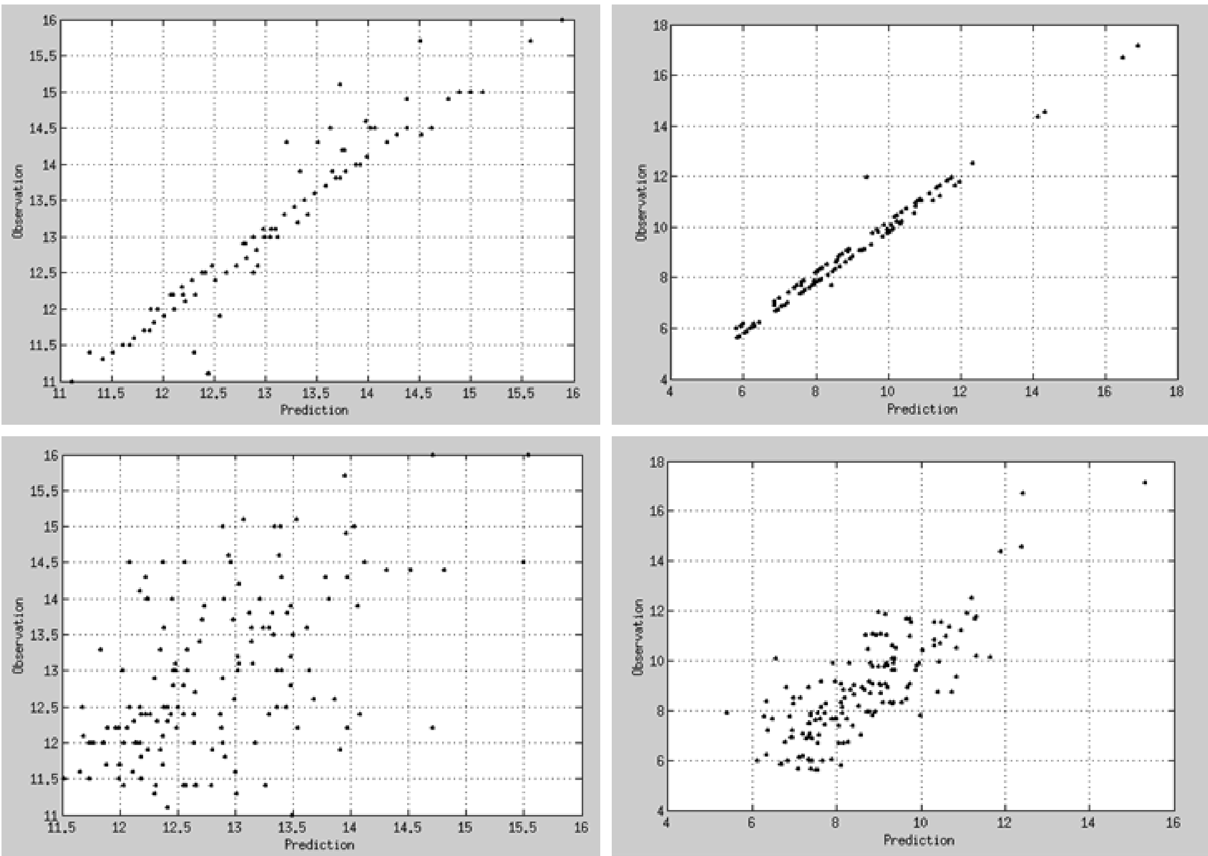

Figure 3: Top graphs represents observation versus prediction for training set. Below of them are located the graphs of the test set. Left are for sugar and right are for acid indexes.

ISSN 0947-7314

4 Conclusions

The reproducibility and internal correlation of measurements made of our newly developed technique is high for both experiments.

After the experiment was done, an accuracy near to twenty per cent of error in sugar indexes and only ten per cent in acid indexes were obtained. This result implies that this method it is enough worthy to know beforehand the sugar and acid of an orange, and that information can be used by postharvest centers to choose the destination of every orange.

Acknowledgements

This work has been partially supported by projects CSD2007-00018 and PET2005-0643 from the Spanish Ministry of Science and Innovation, and by the project P1 1B2007-48 from Fundació Caixa-Castelló.

References

HSU C.-W., CHANG C.-C., LIN C.-J. (2003): A Practical Guide to Support Vector Classification, Technical Report, Department of Computer Science, National Taiwan University. [Online]. Available: www.csie.ntu.edu.tw/_cjlin/libsvm

LI, W., GOOVAERTS P., MEURENS M.(1996): Quantitative analysis of individual sugars ans acids in orange juices by near-infrared spectroscopy of dry extract, Journal of Agricultural and Food Chemistry, 44: 2252-2259

NICOLAÏ B.M., BEULLENS K., BOBELYN E., PEIRSA., SAEYS W., THERON K.I., LAMMERTYN J. (2007): Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review, Postharvest Biology and Technology, 46: 99-118

STEINMETZ V., BIAVATI E., MOLTÓ E., PONS R., FORNES I., (1995): Predicting the maturity of oranges with non destructive sensors, In Acta Hort. 421. Ed.: Th. H. Gieling, B. Bennedsen, W. T. M. van Meurs; 2nd International Symposium on Sensors in Horticulture, 421: 271-278

TSUCHIKAWA S., HAMADA T. (2004): Application of time-of-flight near infrared spectroscopy for detecting sugar and acid contents in apples, Journal of Agricultural and Food Chemistry, 54 (9): 2434-2439