Software instrument for classification of milk products by analysis of spectrophotometric data

Stanislav Penchev

A Training Model of a Microprogramming Unit for Operation Control: *The paper describes the use and the main blocks of a software package for classification of food products, presented by features derived from spectrophotometric characteristics, into quality groups. Methods and tools used for data dimensionality reduction and for classification are described and an example of using the software for classification of spectral characteristics of white cheese in two classes is presented. All the procedures are developed in MATLAB environment.*

Key words: *spectral characteristics, classification, food quality.*

INTRODUCTION

The application of spectroscopic investigations in the near infrared (NIR) are becoming increasingly popular for analysis of the production of animal origin. NIRS allows for measurements in real time (on-line) of parameters in all stages of production – from obtaining of raw materials to the production of the final, ready-to-eat product.

Unlike computer vision, spectral analysis has found wider application in the evaluation of quality and safety of dairy products. The possibilities for application of spectral analysis in the near infrared and mid-infrared region are well studied in the assessment of dairy products. This includes monitoring the production process, determining the geographical origin, quality assessment, monitoring of the maturation process of some dairy products and others.

Surveys on the possibilities of NIRS for assessment of different properties of milk and milk products associated with their quality and safety are published by a number of authors [7, 11]. They state that the analytical methods, including NIRS, combined with appropriate chemometric instruments are powerful tools with great potential for quality assessment and identification of milk and dairy products.

As typical examples of the application of NIRS for assessing various characteristics associated with quality and safety of dairy products can be mentioned:

Determination of the composition and quality of milk. In [11] a study in which the authors apply spectral analysis of milk products in the near infrared region using FT-NIR method for testing yoghurt is presented. In [3, 4, 8] other applications of spectral analysis for assessment of characteristics of milk, associated with its quality and safety like fat content lactose, proteins and others are described.

Determination of the sensory properties and age of the cheese. Sensory properties are important attribute for quality of dairy products since they are directly related to the initial evaluation of users and experts. The evaluation of sensory properties by analysis of spectrophotometric data is subject of examination by various authors [4, 5, 7].

In the [5] the potential of MIR spectroscopy in the wavelength range 640 to 4000 nm, in combination with PLSR is used for predicting the various indicators of the quality and safety of cheddar cheese. Studies related to the capabilities of the NIR spectral analysis to determine the range of various sensory properties of dairy products such as taste (salt content, flavor of oil on rancidity, spice), texture properties (hardness, chewiness and elasticity), cheeses made from cow, sheep and goat milk have been published by other authors [1, 2, 10].

Determination of the composition of cheeses. Comparative analysis of two methods - OT-NIR and FTIR-ATR to predict the fat content, protein and water [7] were applied to 24 samples of cheese. NIR reflectance spectra, combined with chemometric methods is used to determine the level of protein, moisture, fat, salt, peptides and other ingredients for various types of cheeses.

Analysis of the microbiological composition. In a study of the processes of change in pasteurized milk [2] the authors found a correlation between microorganisms present, pH, and the spectral data obtained in the range 600-1100nm. The variety of fungi of the genus Lactococcus in the ripening period of the soft cheese produced from two different starter cultures (SA and SB), is analyzed by FTIR spectroscopy [9]. In [6] NIRS technology is used for the analysis of hydrophilic (HI) and hydrophobic (HO) microorganisms and ratio HO / HI is evaluated.

MATERIALS AND METHODS

Fig. 1 shows the general appearance of the Graphical User Interface (GUI) of the software developed. It is designed to solve a specific task: evaluation of data separability for properties derived from selected frequency bands, on which the spectral range of the device is divided. The task provides aggregation of the neighboring points of the spectral characteristic, in order to obtain hyperspectral characteristic, consisting of various non-overlapping spectral bands. The goal is to find the minimum required number of frequency bands that provides the best separability of data classes.

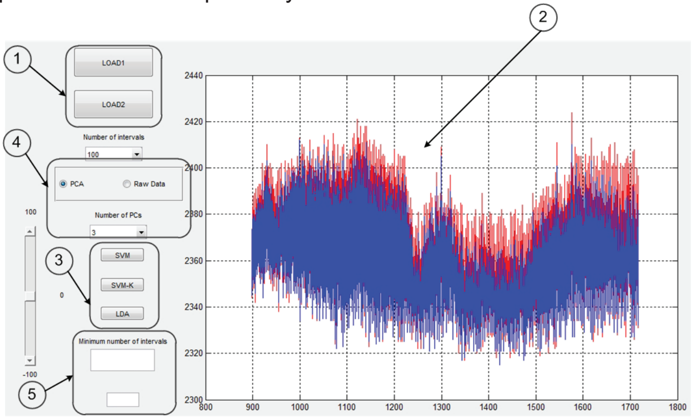

Fig. 1 Graphical user interface of the software.

The graphical interface is composed of several basic blocks. Buttons in block 1 serve to load the spectral characteristics of multiple objects from two classes that are displayed in the graphic field 2. In block 4 a choice of the way of presenting the data is provided: by properties derived using data dimensionality reduction by Principal Component Analysis (PCA) or with the original spectral characteristics. Upon presentation of the data by Principal Components (PC), it is possible to indicate the number of PC's in the range from 3 to 10. By controls in block 3 a choice of a classification procedure is made, which will be used for assessment of data separability. The results of separability assessment are presented in block 5.

MATLAB is used for software development, because of the rich mathematical abilities of that environment and the ability to use a variety of built-in procedures for classification and for assessment of classification accuracy.

Data dimensionality reduction. To extract the properties of spectral characteristics and to reduce the dimensionality of the spectral data principal components analysis (PCA) is used.

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Consider a data matrix, X, with column-wise zero empirical mean, where each of the n rows represents a different repetition of the experiment, and each of the *p* columns gives a particular kind of datum. Mathematically, the transformation is defined by a set of p-dimensional vectors of weights or loadings $w_{(k)} = (\varpi_1, \dots \varpi_p)_{(k)}$ that map each row $x_{(i)}$ of X to a new vector of principal component scores $t_{(i)} = (t_1, \dots, t_p)_{(i)}$, given by

$$t_{k(i)} = x_{(i)} \cdot w_{(k)} \tag{1}$$

in such a way that the individual variables of t considered over the data set successively inherit the maximum possible variance from x, with each loading vector w constrained to be a unit vector.

For representation of the spectral characteristics the first few principal components are used, corresponding to the frequencies with the largest variance, as in the software that number can be set in advance, and ranges from 3 to 10.

Classification procedures. Separability of data for the same area of the product on different days of storage is a major criterion for the correct classification both with regard to composition of the product and of its freshness. The separability between the data of the investigated areas was quantitatively assessed by the overlap error ε_{pr}% (the ratio of incorrectly classified examples to the total number of examples). The separability determination was made by three methods: LDA, which implements linear separability, linear SVM and kernel version of SVM, which satisfies the conditions for linear separability of the classes of data.

EXPERIMENTAL RESULTS.

Multiple spectral characteristics of different food products in the visible and near infrared region are preliminary recorded, in order to trace their change during storage of food samples in irregular operating conditions (at a temperature of 20ºC and lack of light). The products used are meat, bacon, white cheese and yellow cheese. Data is structured in such a way that allows solving the following several tasks: separability assessment of individual samples on a certain day of the storage of the product and separability assessment of data for the same sample on different days of storage.

The study provides dividing the entire spectral range to different number of intervals and to find the minimum number, which ensure the best separability of data classes. For this purpose, the characteristic is divided successively at intervals of 2 to 100. Classification in two classes, represented by the way chosen in block 4 (Fig. 1) and using the classifier chosen in block 3, is performed for every interval. The class separability is assessed quantitatively by the ratio of incorrectly classified examples to the total number of examples ε_{pr}% for each of the intervals. Finally, as a result the behavior of this classification error is shown graphically. The number of intervals and the number of the interval with the least error are also shown.

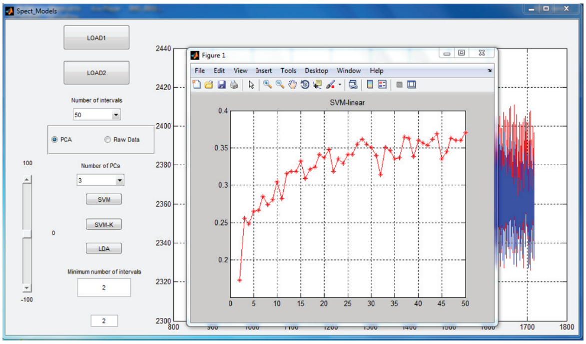

Fig. 2 Results when using linear SVM classifier.

As an example of the operation of the program the classification results of samples of white cheese on the first day (class1) and on the fourth day (class2) of storage are shown. Their spectral characteristics are taken in near infrared region (900 ÷ 1700nm), and represented by its first three PC's. Fig. 2 to 4 present the results of the program work using a linear SVM classifier (Fig. 2), kernel-SVM classifier (Fig. 3) and LDA classifier (Fig. 4).

The investigation when using the SVM classifier shows that the error reaches its minimum (17.3%) when the characteristic is divided in 2 intervals (Fig. 2). The interval in which that error has been reached is number 2.

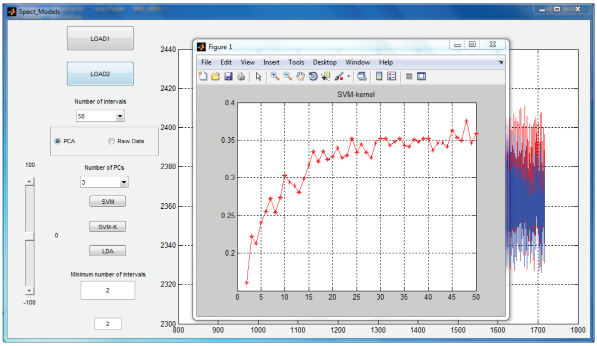

Fig. 3 Results when using kernel SVM classifier.

When using kernel SVM classifier (Fig. 3) best separability (ε_{pr}=16.1%) is reached in dividing the characteristic into 2 intervals. The interval in which that error has been reached is number 2.

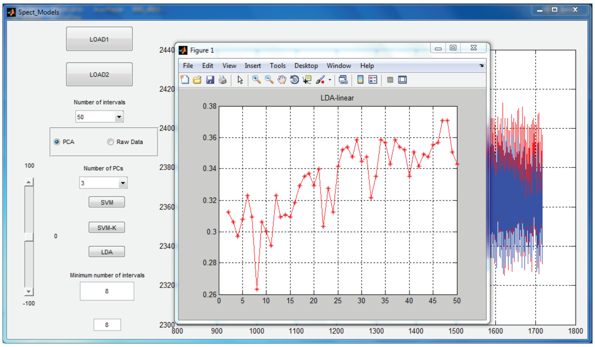

Fig. 4 Results when using LDA classifier.

The use of LDA classifier (Fig. 4) show, that the best separability (ε_{pr}=26.3%) is reached in dividing the characteristic into 8 intervals. The interval in which that error has been reached is number 8.

CONCLUSION

The performed research and presented results allow to draw the following main conclusions:

1. The developed software package for the classification of food products by spectrophotometric data allows different investigations, concerning both the classification procedures and the presentation of the input data, to be performed. The approach used in the software package allows assessment of changes in product quality during storage, the change in the quality of individual areas and recognition of these areas within the same sample.

2. The results obtained show that the approach used to separate the spectral characteristics in intervals and to search for the most informative parts thereof, has a potential for solving the problems with reasonable accuracy both in the visible and near-infrared spectral range.

3. The embedded in the software approach for searching the most informative parts of the spectral characteristic allow significant acceleration of the evaluation process.

REFERENCES

[1] Alander, J. T, V. Bochko, B. Martinkauppi, S. Saranwong, T. Mantere. A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. International Journal of Spectroscopy, Volume 2013, 2013.

[2] Al-Qadiri H. M. Lin M. Al-Holy M. A., Cavinato A. G. and Rasco B. A. Monitoring Quality Loss of Pasteurized Skim Milk Using Visible and Short Wavelength Near-Infrared Spectroscopy and Multivariate Analysis J. Dairy Sci. 91, 2008, pp.950–958.

[3] Da-Wen Sun (editor), Infrared Spectroscopy for Food Quality Analysis and Control, Academic Press / Elsevier, San Diego, California, USA, 2009.

[4] Drackova, M., L. Hadra, B. Janstova, P. Havratilova, H. Pridalova, L. Vorlova. Analysis of goat milk by near-infrared spectroscopy. Acta vet. Brno 2008.

[5] Fagan, C. C., O'Donnell, C. P., O'Callaghan, D. J., Downey, G., Sheehan, E. M., Delahunty, C. M., Everard, C., Guinee, T. P. and Howard, V. Application of mid-infrared spectroscopy to the prediction of maturity and sensory texture attributes of cheddar cheese. Journal of Food Science, 72 (3), 2007, pp. 130-137.

[6] Gonzalez-Martin, I., Hernandez-Hierro, J. M., Vivar-Quintana, A., Revilla, I., Gonzalez-Perez.The application of near infrared spectroscopy technology and a remote reflectance fibre-optic probe for the determination of peptides in cheeses (cow’s, ewe’s and goat’s) with different ripening times. Food Chem 114, 2009, pp. 1564-1569.

[7] Gunasekaran, S., M. Mehmet Ak. Cheese Rheology and Texture [Kindle Edition]. CRC Press; 1 edition, 2002.

[8] Laporte, M., P. Paquin. Near-infrared analysis of fat, protein, and casein in cow's milk, Journal of Agricultural and Food Chemistry, vol. 47, no. 7, 1999, pp. 2600–2605.

[9] Lefier, D., Lamprell, H. and Mazerolles, G. Evolution of Lactococcus strain during ripening in Brie cheese using Fourier transform infrared spectroscopy. Lait., 80, 2000, pp. 247-254.

[10] Meagher, L.P., Holroyd, S.E., Illingworth D., van de Ven F. and Lane S. At-line near infrared spectroscopy for prediction of solid fat content of milk fat from New Zealand butter. J. Agri. Food Chem., 55, 2007, pp. 2791-2796.

[11] Růžičková, J., K. Šustová. Determination of selected parameters of quality of the dairy products by NIR spectroscopy. Czech J. Food Sci., 24, pp. 255–260, 2006.

За контакти:

гл.ас. д-р Станислав Пенчев, Катедра “Автоматика и мехатроника”, Русенски университет “Ангел Кънчев”, тел.: 082-888 379, e-mail: msp@uni-ruse.bg

Докладът е рецензиран.