Journal of Food Engineering 149 (2015) 38–43

Contents lists available at ScienceDirect

Journal of Food Engineering

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

E-nose combined with chemometrics to trace tomato-juice quality

Xuezhen Hong [a,b], Jun Wang [a,*], Guande Qi [c]

[a] Department of Biosystems Engineering, Zhejiang University, 688 Yuhangtang Road, Hangzhou 310058, PR China
[b] College of Quality & Safety Engineering, China Jiliang University, Hangzhou 310018, PR China
[c] Department of Computer Science, Zhejiang University, Hangzhou 310027, PR China

ARTICLE INFO

Article history:
Received 20 June 2014
Received in revised form 17 September 2014
Accepted 1 October 2014
Available online 12 October 2014

Keywords:
Electronic nose
Cherry tomato juice
Storage
Trace quality
Semi-supervised classification
Partial least squares regression

ABSTRACT

An e-nose was presented to trace freshness of cherry tomatoes that were squeezed for juice consumption. Four supervised approaches (linear discriminant analysis, quadratic discriminant analysis, support vector machines and back propagation neural network) and one semi-supervised approach (Cluster-then-Label) were applied to classify the juices, and the semi-supervised classifier outperformed the supervised approaches. Meanwhile, quality indices of the tomatoes (storage time, pH, soluble solids content (SSC), Vitamin C (VC) and firmness) were predicted by partial least squares regression (PLSR). Two sizes of training sets (20% and 70% of the whole dataset, respectively) were considered, and $R^2 > 0.737$ for all quality indices in both cases, suggesting it is possible to trace fruit quality through detecting the squeezed juices. However, PLSR models trained by the small dataset were not very good. Thus, our next plan is to explore semi-supervised regression methods for regression cases where only a few experimental data are available.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The quality of fruit is relatively easy to identify by their morphological characteristics (such as color, texture and firmness) and flavor (odor and taste). However, the act of processing fruits into juices makes the quality difficult to identify, e.g. it is hard to tell whether a bottle of juice is squeezed from fresh fruits or not. Therefore, it is important to develop a method that could identify and trace quality of raw fruits by detecting the squeezed juices.

E-nose, which has been inspired by the way human recognize samples via olfaction, has proven to be a good tool for quality assessment (Gil-Sánchez et al., 2011). A typical e-nose system contains a non-selective sensor array, a signal processing subsystem and a pattern recognition subsystem (Gardner and Bartlett, 1994; Wang et al., 1997; Winquist et al., 1997). In the area of fruits and juices detection, the e-nose has been applied for early detection of *Alicyclobacillus* spp. in peach, orange and apple juices (Gobbi et al., 2010), classification of citrus juices according to fruit type (Reinhard et al., 2008), detection of orange juice treatments (Shaw et al., 2000), prediction of peach quality indices (Zhang et al., 2012), authentication of cherry tomato juices (Hong et al., 2014b), and monitoring of fruit ripeness (Brezmes et al., 2000), etc. In the abovementioned and many other applications, supervised classification is a fundamental data analysis task (Scott et al., 2006). A lot of supervised classification methods, e.g. linear discriminant analysis (LDA) (Vera et al., 2011), quadratic discriminant analysis (QDA) (Cerrato Oliveros et al., 2002), classification and regression trees (CART) (Buratti et al., 2004), classification and influence matrix analysis (CAIMAN) (Ballabio et al., 2006), various neural networks (NNs) (Hong et al., 2012), support vector machines (SVM) (Brudzewski et al., 2004) and random forest (RF) (Pardo and Sberveglieri, 2008), have been successfully applied for e-nose data analysis. Generally, supervised classification requires sufficient labeled data to train a good classifier (sufficient usually means that the labeled data can roughly represent the underlying structure of the entire data space) (Gan et al., 2012). If the labeled data only represent part of the underlying data structure, or if the labeled data are mostly consisted of outliers, the classifier built would lack generalization, i.e. it cannot function well for the testing data. However, sufficient labeled data requires extensive effort, money, materials and time. Therefore, it is important to find a classification approach that can perform well with fewer labeled training data.

Semi-supervised classification, which uses unlabeled data together with labeled data to build a better classifier, has become a recent topic of interest especially in the area of computational statistics (Vandewalle et al., 2013), image analysis (Filipovych and Davatzikos, 2011), network traffic (Erman et al., 2007), document classification (Shi et al., 2011) and biomedical informatics

* Corresponding author. Tel.: +86 571 88982178; fax: +86 571 88982191.
E-mail address: jwang@zju.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.jfoodeng.2014.10.003
0260-8774/© 2014 Elsevier Ltd. All rights reserved.

(Garla et al., 2013), etc. Our research is inspired by Cluster-then-Label – a semi-supervised approach based on clustering (Zhu and Goldberg, 2009). The main advantage of this approach is that it reveals the actual data space structure through clustering both labeled and unlabeled data to compensate for the limitation of labeled data. It is noticeable that this clustering based semi-supervised approach is very sensitive to its underlying assumptions, i.e. the clusters coincide with decision boundaries. If the assumption is incorrect, the results would be poor. However, the clustering algorithms that are mostly applied in the area of e-nose have their own limitations and scopes of application (Hong et al., 2014a). Thus, in this paper, a state-of-the-art clustering approach – spectral clustering – is also introduced. By constructing an undirected weighted similarity graph on the data, spectral clustering utilizes the spectrum of the graph Laplacian to obtain a low dimensional representation of the data, and then does clustering using classical methods, such as *k*-means (Chen et al., 2011). This graph-theoretic based clustering method can be solved efficiently by standard linear algebra software and very often outperforms conventional clustering algorithms (Von Luxburg, 2007).

In this paper, a PEN 2 e-nose was applied to identify freshness of *youbei* cherry tomatoes that were squeezed for juice consumption. A Cluster-then-Label approach based on spectral clustering and majority voting was applied to deal with the e-nose data for the first time. Classifications of the e-nose dataset by the semi-supervised approach and various supervised approaches were compared. The main objectives of this research are: (1) to explore if it is possible to trace fruit quality by detecting the squeezed fruit juice using the e-nose technique, and (2) to explore if the proposed semi-supervised approach would outperform supervised approaches in the case that only a few labeled e-nose data was available for training.

2. Experimental

2.1. Preparation of tomato juice samples

Chinese variety, *youbei* cherry tomatoes were hand harvested three times (every 6 h) from different orchards located at Hangzhou, China. All tomatoes were picked at light red stage (approximately 70% of the surface, in the aggregate, shows pinkish-red or red) (Agriculture, 1997). Upon arrival at the laboratory, the cherry tomatoes were selected according to approximately uniform size and weight and non-damaged and not-attacked by worm. Selected samples were then rinsed with clear water and wiped dry with clean cloth prior to being stored in a refrigerator at 4 °C for 16 days.

The e-nose measurements were conducted every three days, i.e. on day 1, 4, 7, 10, 13 and 16. On each measuring day, appropriate amount of cherry tomatoes were placed in a fruit squeezer and juiced for 30 s to obtain 100% fresh juices. The juicing process was repeated 25 times. Thus, there were in total 150 (25 samples × 6 storage time (ST)) juice samples for the e-nose detection.

2.2. E-nose and sampling procedure

The headspace analysis was performed with a commercial PEN 2 e-nose (Airsense Analytics, GmBH, Schwerin, Germany) containing ten metal-oxide semiconductors. Description of the sensor array has been given in our previous works (Hong et al., 2012).

Prior to detection, each sample (10 mL of cherry tomato juice) was placed in a 500 mL airtight glass vial that was sealed with plastic wrap for 10 min (headspace-generation time). During the measurement process, the headspace gaseous compounds were pumped into the sensor arrays through Teflon tubing connected to a needle in the plastic wrap, causing the ratio of conductance G/G_0 (G and G_0 are conductance of the sensors exposed to sample gas and zero gas, respectively) of each sensor changed. The measurement phase lasted for 70 s, which was long enough for the sensors to reach stable signal values. The signal data from the sensors were collected by the computer once per second during the measurements. When the measurement process was complete, the acquired data were stored for later analysis. After each measurement, zero gas (air filtered by active carbon) was pumped into the sample gas path from the other port of the instrument for 60 s (flush time). In case of sensor pollution which could cause sensor drift, after all the measurements were done, nitrogen gas was pumped into the sample gas path to clear the sensor array.

2.3. Measurements of pH, SSC, VC and firmness

On each measuring day, pH, soluble solids content (SSC, ^{0}Brix), Vitamin C (VC) and firmness of the cherry tomatoes were also measured. For each quality index, 25 replicates were prepared (to correlate with the numbers of e-nose/e-tongue measurements). pH was measured by a titrimeter (Ti-Touch-916, Metrohm, Switzerland). SSC was measured by a temperature compensating refractometer in ^{0}Brix (Digital refractometer 2WA-J 0–32% Shanghai, China). VC concentration was measured by the method of 2,6-dichloro-indophenol titration according to National Standard of the People's Republic of China (GB/T 6195-1986, 1986), and its value was expressed as mg ascorbic acid per 100 g of tomato (mg/100 g). Puncture process that was measured using a Universal Testing Machine (Model 5543 Single Column, Instron Corp., Canton MA, USA) was expressed as cherry tomato firmness. The penetrating force of all individual fruit was measured on the three positions along the equator approximately 120° between them, perpendicular to the stem-bottom axis. A 6 mm diameter stainless steel cylindrical probe with a flat end was used. The puncture process was auto recorded by computer, and the final puncture force was defined as the average of three maximum forces required to push the probe to a depth of 3 mm at a speed of 5 mm s^{-1}.

All the experiments and measurements were carried out at a room temperature of 25 ± 1 °C.

3. Data analysis methods

3.1. Semi-supervised approach

As the name implies, Cluster-then-Label is a semi-supervised classification approach of incorporating label information into unsupervised clustering. The algorithm applied in this paper proceeds as follows:

Input: labeled and unlabeled data, un-supervised clustering algorithm A, and a supervised learning algorithm L
(1) Cluster all the data (labeled and unlabeled) using A.
(2) For each resulting cluster, let S be the labeled instances in this cluster. Then, learn a supervised predictor from S: fs = L(s), and apply fs to all unlabeled instances in this cluster.
Output: labels on unlabeled data.

The supervised learner L used here is majority voting – each cluster is assigned a class label corresponding to the majority class of points belonging to that cluster. Actually, one of the cluster validation criteria – precision-recall measure – is also based on majority voting. The unsupervised clustering algorithm A applied here is spectral clustering. Description of this clustering algorithm could be found in our previous paper (Hong et al., 2014a).

Download English Version:

https://daneshyari.com/en/article/222952

Download Persian Version:

https://daneshyari.com/article/222952

Daneshyari.com