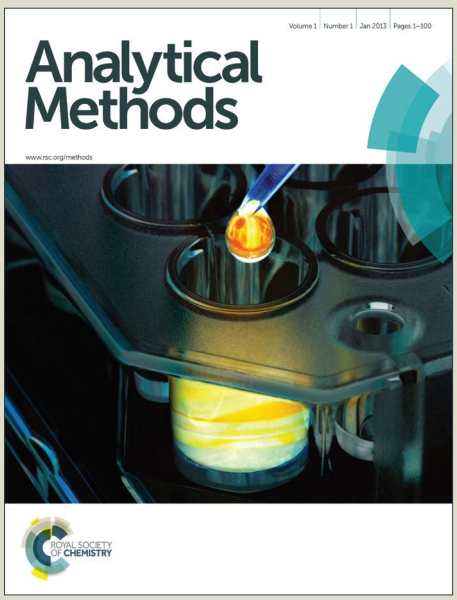
Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Rapid Identification between Edible Oil and Swill-Cooked Dirty Oil by Using Near-infrared Spectroscopy and Sparse Representation Classification

Yang Zhou,[a,b] Tiebing Liu[a], Jinrong Li[a] and Zhengwei Chen[a]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

Rapid identification between edible oil and swill-cooked dirty oil was a challenging and significant task in the field of food safety. The main object of this investigation was to identify between edible oil (QO) and swill-cooked dirty oil (SO) by employing near-infrared (NIR) spectroscopy and sparse representation classification method (SRC). For the diversity and uncertainty of the species of swill-cooked dirty oil, building classification model based on NIR spectroscopy faced the problems of complex system and small sample. In order to improve the stability and accuracy of the identification task, the SRC trained the redundant dictionaries for QO and SO, and figured out the sparse representation coefficients of spectroscopy in validation set under both dictionaries. Then the spectroscopy in validation set was reconstructed by these sparse representation coefficients and its corresponding dictionary. Finally, the reconstruction errors under QO and SO dictionaries were used as indicator for classification. Moreover, an simplified SRC algorithm (SRC-S) which directly used the spectrum of calibration set as redundant dictionaries was proposed, and it got rid of dictionary training process that avoided the information loss during the training. Comparing with the linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA), the experiment results showed that the SRC-S outperformed the SRC, and it reached the best classification accuracy of 95.37%, which proved it was possible to identify between QO and SO by using NIR spectroscopy and SRC-S.

Introduction

Swill-cooked dirty oil is contaminated by bacteria, heavy metals and harmful chemical composition during the process of collecting, refining and storing, and it also produces a variety of carcinogenic substances in the recycling and store procedure, which seriously harms public health and becomes a hotpot in food quality and safety.[1,2] Acid value, relative unsaturated degree of fatty acid, aldehyde/ketone volatile composition content, solid fat content, cholesterol content, heavy metals content and electrical conductivity are used as the traditional method for edible oil quality analysis, but these conventional physical and chemical indicators have been unable to finish the task of identification between edible oil and swill-cooked dirty oil after several procedures, such as heating, decoloring, deacidification and thickening.[3] Therefore, it is significant and critical to develop a rapid and effective method for swill-cooked dirty oil identification.

Due to the sensitivity of organic compounds, the near infrared (NIR) spectroscopy has gradually used for oil identification in recent years. Widespread brands, registered designation of origin, and cultivar of edible oil are classified by these techniques.[4-6] Many kinds of vegetable oil from different manufacturers, namely soybean oil, palm oil, sesame oil, peanut oil and so on, are distinguished by using a Fourier transform NIR spectrometer.[7] The research results show that the NIR technique has the ability to distinguish different kinds of oils with advantages of fast speed, non-destructive, low cost and non-pollution, which makes the identification between edible oil and swill-cooked dirty oil feasible.

At present the domestic and foreign research of NIR technology mainly focuses on identifying a specific variety of oil from different unknown varieties through near infrared spectrum. But in practical application, edible oil may consist of colza oil, soybean oil, sesame oil, blend oil, etc.[8] Swill-cooked dirty oil can be collected from the sewer floater, oil slick of hogwash, or made from animal meat, internal organs and skin, or recycle from waste oil, even made by the mixture of gutter oil and edible oil.[9] Because of the diversity and uncertainty of the species of swill-cooked dirty oil and edible oil, the spectra are multicollinearity and overlapping, so building spectral quantitative model meets some problems such as complex system and small sample,[10] which increase the difficulty of the swill-cooked dirty oil authentication. In such a small and complex sample system, the conventional clustering methods, like K-means clustering, fuzzy k-means clustering and SIMCA, which were used in NIR spectroscopy analysis, can't meet the needs of swill-cooked dirty oil identification.

With the development of sparse theory, the sparse features of spectral signal are used for classification, which obtain better performance than traditional classification method.[11,12] The sparse representation classification method extracts the feature in face image in which it has made the preliminary investigation,[13,14] and gradually extends to other areas. Sparse representation classification is an effective classification method in various fields, but these methods have not been

reported in the application of near infrared spectral classification.

In this paper, we employ a sparse representation classification method (SRC) for identification between edible oil and swill-cooked dirty oil based on NIR spectroscopy. Its main advantage is that different kinds of edible oil and swill-cooked dirty oil samples are used to train its own redundant dictionaries. The spectral sparse representation coefficients of validation sample under different trained redundant dictionaries are calculated, then these sparse representation coefficients, combining with trained redundant dictionaries, are used for spectral reconstruction. Finally, the validation sample is classified according to minimum sparse representation error under different trained redundant dictionaries. Because of the complexity of the dictionary training, we improve the SRC method. Its simplified algorithm (SRC-S) is applied for swill-cooked dirty oil identification in which the SRC-S directly uses the calibration samples as the redundant dictionaries.

Spectral samples of various edible oil and swill-cooked dirty oil were collected in experiment, and the linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA) were used as contrastive approach. Comparing with these methods, the results show the SRC-S had a global improvement, which proved that NIR technology with Chemometrics of SRC-S was feasible to rapidly distinguish between edible oil and swill-cooked dirty oil.

Materials and methods

Samples

The oil samples used as calibration and validation set were provided by Quality Inspection Institute of ShaoXing City, China, in 2013. A total of 168 oil samples included 60 qualified edible oil samples and 108 swill-cooked dirty oil samples. The edible oil samples consisted of colza oil, soybean oil, sesame oil, peanut oil and sunflower oil, and each kind had 12 samples. By the end of sample collecting, there was no standard of swill-cooked dirty oil authentication and all swill-cooked dirty oil samples were ferreted out by inspectors.

Acquisition of the near infrared spectrum

The near infrared spectroscopy of all samples were recorded on a Fourier transform (FT-NIR) spectrometer (MPA Bruker) over the 12500 cm^{-1} to 4000 cm^{-1} frequency range at a resolution of 4 cm^{-1}, which equipped with a liquid optical fiber attachment set at 2mm optical distance. The NIR spectrum of oil samples is shown in Figure 1. Each sample was scanned 8 times and the average spectroscopy was used for calibration and validation process.

Software and computing

The data processing algorithm was programmed on the Matlab 7.0 software with PLS Toolbox. The initial spectra were collected by OPUS 6.5 and then it was changed to the 1×2051 data vectors which was defined by the number of spectral variables. The mean centering and variance normalization were used as pretreatment method to reduce the absorbance discrepancy of oil samples at different wavelength in follow-up experiments.

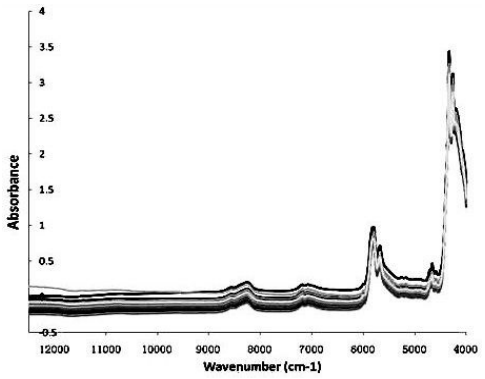


Fig.1. Spectrum of edible oil and swill-cooked dirty oil

Sparse representation classification (SRC)

Sparse classification method was based on sparse representation and each NIR spectroscopy vector x can be decomposed under the redundant dictionary D with size of $n \times m$. The n was the dimension of spectroscopy vector x and the m was the size of sparse representation coefficients α, where the majority of elements in vector α were equal to zero. The amount of non-zero elements in vector α was named as sparsity. This characteristic of α was considered as sparse feature and dictionary D was redundant.

$$x = D\alpha \qquad D \in n \times m \tag{1}$$

In fact, it can't be determined that which redundant dictionary was suit for identification between edible oil and swill-cooked dirty oil. Therefore, spectral samples in calibration set were applied to train the redundant dictionaris with K-SVD dictionary training algorithm which took full account of oil spectral characteristic.[15] The K-SVD dictionary learning algorithm had two stages of batch operation. The first stage was to calculate sparse coefficients of redundant dictionary, and the second stage updated the dictionary, while all spectrum of samples played a part in training at both stages.

After the redundant dictionaries were trained, directly solving the sparse representation coefficients α belonged to NP-Hard puzzle in theory. Therefore, sub-prime approximation method was exploited for solving this problem. In flowing experiment, the sparse representation coefficients α were calculated by orthogonal matching pursuit algorithm (OMP) which was a widely used greedy algorithm to search the second-best solution of α .[16]

The sparse representation classification (SRC) algorithm established redundant dictionaries from both edible oil and swill-cooked dirty oil samples in calibration set. The procedure of solving the sparse representation coefficients was equal to searching spectral decomposition of validation samples under the both dictionaries. Finally, the spectrum of validation samples were reconstituted by different redundant dictionaries and its representation coefficients. The key to SRC was reconstruction errors which stood for similarity between

spectroscopy and different dictionaries,[17-18] and it can be used for predicting the oil sample was swill-cooked dirty oil or not. Based on this idea, the main steps of SRC were described as follows.

(1) The spectrum of oil samples in calibration set were divided into two classes, qualified edible oils (QO) X_{QO} and swill-cooked dirty oil (SO) X_{SO}. Their corresponding redundant dictionaries D_{QO} and D_{SO} were trained by K-SVD algorithm.

(2) One spectroscopy y was extracted from the validation set, and its sparse representation coefficients α_{QO} and α_{SO} under dictionary D_{QO} and D_{SO} were calculated by solving optimization problems. After designating the sparsity (the amount of none-zero elements in α_{QO} and α_{SO}), OMP algorithm was used to solve those optimization problems.

$$\alpha_{QO}^{y} = \arg\min_{\alpha} \|\alpha_{QO}^{y}\|_0 \quad \text{s.t} \quad \|y - D_{QO}\alpha_{QO}^{y}\| < \varepsilon \tag{2}$$

$$\alpha_{SO}^{y} = \arg\min_{\alpha} \|\alpha_{SO}^{y}\|_0 \quad \text{s.t} \quad \|y - D_{SO}\alpha_{SO}^{y}\| < \varepsilon \tag{3}$$

(3) The reconstruction errors of spectroscopy y under dictionary D_{QO} and D_{SO} were figured out and the class label of y was predicted according to decision function listed in equation 6.

$$e_{QO} = \|y - D_{QO}\alpha_{QO}\| \tag{4}$$

$$e_{SO} = \|y - D_{SO}\alpha_{SO}\| \tag{5}$$

$$f(y) = \arg\min\{e_j, j = QO \text{ or } SO\} \tag{6}$$

(4) Above steps were repeated until all the spectrum in validation set were predicted.

Simplified SRC algorithm (SRC-S)

When the amount of the calibration samples was too large, it was time-consuming to finish each dictionary training process and also prematurely stopping iteration reduced the accuracy of the sparse representation. For this reason, we put forward the simplified SRC(SRC-S) algorithm which directly used the spectrum in calibration set as redundant dictionary, thereby it significantly decreased computational complexity. The main steps of SRC were described as follows.

(1) The spectrum of oil samples in calibration set were divided into two classes, qualified edible oils (QO) X_{QO} and swill-cooked dirty oil (SO) X_{SO} which were regarded as dictionaries.

(2) One spectroscopy y was extracted from the validation set, and the sparse representation coefficients α_{QO} and α_{SO} under dictionary X_{QO} and X_{SO} were calculated by solving optimization problems. After designating the sparsity (the amount of none-zero elements in α_{QO} and α_{SO}), OMP algorithm was used to solve those optimization problems.

$$\alpha_{QO}^{y} = \arg\min_{\alpha} \|\alpha_{QO}^{y}\|_0 \quad \text{s.t} \quad \|y - X_{QO}\alpha_{QO}^{y}\| < \varepsilon \tag{7}$$

$$\alpha_{SO}^{y} = \arg\min_{\alpha} \|\alpha_{SO}^{y}\|_0 \quad \text{s.t} \quad \|y - X_{SO}\alpha_{SO}^{y}\| < \varepsilon \tag{8}$$

(3) The reconstruction errors of spectroscopy y under dictionary X_{QO} and X_{SO} were figured out and the class label of y was predicted according to decision function listed in equation 11.

$$e_{QO} = \|y - X_{QO}\alpha_{QO}\| \tag{9}$$

$$e_{SO} = \|y - X_{SO}\alpha_{SO}\| \tag{10}$$

$$f(y) = \arg\min\{e_j, j = QO \text{ or } SO\} \tag{11}$$

(4) Above steps were repeated until all the spectrum in validation set were predicted.

Comparing with original SRC algorithm, SRC-S algorithm retrenched the procedure of dictionary training, whose execution time was far less than that of SRC. It improved the execution efficiency. In the experiment, the classification performance of these two algorithms had been investigated.

Results and Discussion

During the establishing classification model, 60 oils samples from both edible oil (QO) and swill-cooked dirty oil (SO) were randomly employed for training process. The remaining 108 oils samples were recorded as validation set which were predicted by SRC and SRC-S algorithm applet based on Matlab software. The oil samples in calibration and validation set were shown in table 1. The redundant dictionaries D_{QO} and D_{SO} required in SRC algorithm were trained by K-SVD algorithm and the sparse representation coefficients α_{QO} and α_{SO} under both dictionaries were solved by OMP algorithm.

In the calculation process, reasonable iteration stop condition, the amount of none-zero elements, was the key factor of prediction. It was inevitable that the NIR spectrum were contaminated by noise during the spectral collection. The sparse noise reduction theory proved that proper sparsity in the sparse representation coefficients should be focused because of the uncertainty of the noise in the spectrum.[19] The average classification errors of SRC and SRC-S within the sparsity range 1 to 20 in OMP were observed.

Tab.1 Oil samples in calibration and validation set

set	class	number of samples	type
calibration	QO	24	colza oil, soybean oil, sesame oil, peanut oil and sunflower oil
	SO	36	seized
validation	QO	36	colza oil, soybean oil, sesame oil, peanut oil and sunflower oil
	SO	72	seized

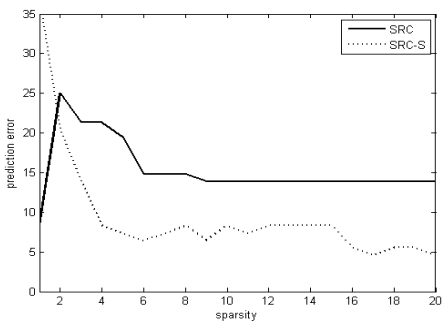


Fig.2. SRC and SRC-S prediction result at different sparsity

Figure 2 shows the prediction results of validation set. The prediction errors of both methods decrease along with the increasing of sparsity. The SRC reaches the optimal average prediction error when the sparsity is equal or greater than 9, whereas that of SRC-S is equal to 17. As a whole, it is easy to find that the error curve of SRC-S is blew that of SRC, which indicates the classification performance of SRC-S is better. We notice that the optimal average prediction error of SRC is 13.89%, while that of SRC-S is 4.63%.

We originally thought that the changing of sparsity would have achieved superior classification accuracy of SRC. Then we tried to use different sparsity and carried on the prediction many times, but the classification results still didn't improve. Due to insufficient of calibration samples, the training process of redundant dictionary caused the loss of class information. However, at the beginning of application, there were not enough calibration samples which can't meet the demand of dictionary training. Also dimension of oil NIR spectrum was 2051 which was far greater than the amount of calibration samples that led to compression of class information. The SRC-S, omitting the step of dictionary training, avoided the information loss and completed task of identification between edible oil and swill-cooked dirty oil. With such complex system and small sample problems, SCR-S overcame multi-collinearity and overlapping of the spectrum and built a spectral quantitative model that was worth being promoted.

Discriminant classification methods, such as linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA) whose aim were to build a qualitative model, represented important tools for solving spectral problems in several areas of oil classification based on NIR spectroscopy.[20,21] Those methods partitioned the multidimensional space of the NIR variables according to the training set, then samples in validation set within the region of space were associated to particular classes. At last, these samples were uniquely assigned class labels. The LDA and PLS-DA were used as the traditional chemometrics for comparison, and the experimental results of LDA, PLS-DA, SRC and SRC-S were reported in Table 2.

Tab.2 Comparison of LDA, PLS-DA, SRC and SRC-S methods

methods	sensitivity	specificity	accuracy
LDA	72.22%	83.33%	79.63%
PLS-DA	77.78%	86.11%	83.33%
SRC	58.33%	100%	86.11%
SRC-S	97.22%	94.44%	95.37%

Table 2 shows that the SRC is slightly better than LDA and PLS-DA. The sensitivity, specificity and classification accuracy of SRC-S are 97.22%, 94.44% and 95.37%, which gets the best performance in identification between QO and SO. The LDA assumed that the sample belonged to a particular class under the hypotheses of the hyper-surfaces separating, and required the explicit calculation of this hyper-surfaces. But the ingredients of SO were similar to QO, it was very hard to describe hyper-surfaces with limited samples in each oil type. The regions in multidimensional space were also partly overlapping. As discussed above, LDA suffers from limitation of uncertainty of SO type. The PLS-DA tries to overcome these limitations by using latent (abstract) variables instead of spectrum experimentally measured. When spectroscopy was projected onto a set of latent variables, such as orthogonal (i.e., not correlated) axes, obtaining the representation of the samples in a subspace was sensitive to oil samples itself. Describing it differently, there was no standard for SO authentication which restricted the stability of the PLS-DA model. Nevertheless, the SRC-S constructed the dictionaries by calibration spectrum, and directly decomposed the spectroscopy in validation set under those dictionaries which took the features from all samples into account simultaneously and was suitable for complex system and small sample spectral analysis.

It is possible to observe lots of noise at the regions of 12500-9000 cm^{-1} in figure 1. The poor signal-to-noise ratio is considered as the reason of poor performance of classification methods, so these regions are discarded and the spectrum of rest regions of 9000-4000 cm^{-1} are used for identification. The experimental results of LDA, PLS-DA, SRC and SRC-S dealing with rest regions are reported in Table 3.

Tab.3 Prediction results at rest regions

methods	sensitivity	specificity	accuracy
LDA	77.78%	83.33%	81.48%
PLS-DA	83.33%	87.50%	86.11%
SRC	63.89%	100%	87.96%
SRC-S	97.22%	94.44%	95.37%

When the regions with poor signal-to-noise ratio were discarded, the performance of LDA, PLS-DA and SRC was slightly improved. But the performance of SRC-S had not changed anymore. According to the sparse representation theory, the procedure of sparse decomposition and reconstruction implied the de-noising. The de-noising pretreatments like smoothing didn't apply before the classification in order to observe the noise impact on different methods. It was obvious that the noise in oil spectrum lower the prediction accuracy. However, the noise was not the key factor to identification between edible oil and swill-cooked dirty oil because of slightly improving after removing the poor regions. Meanwhile the performance of LDA, PLS-DA and SRC-S was slightly improved. Whereas the SRC owned the optimal performance which didn't change whether the regions with poor signal-to-noise ratio were discarded. This indicated that the SRC-S had ability to reduce the noise during sparse decomposition and reconstruction. Also the SRC had the same procedure, but the noise also interfered the calculation of classification. The training of dictionary needed large amount of

samples, and the lack of calibration samples reduced the stability of prediction. In summary, comparing with the traditional methods like LDA and PLS-DA, the experiment results verified that the NIR spectroscopy with SRC-S was feasible and superior for rapid identification between edible oil and swill-cooked dirty oil.

Conclusions

We put forward a framework for identification between edible oil and swill-cooked dirty oil by using SRC-S and NIR spectroscopy. The main advantage of SRC-S was that the sparse representation under redundant dictionary revealed more information and enhanced the spectral resolution in complex and small sample system. The investigation was summarized as follows:

(1) A new simplified SRC-S algorithm was proposed to identify between edible oil and swill-cooked dirty oil by using NIR spectroscopy.

(2) Comparing with SRC, the SRC-S took the calibration samples as redundant dictionaries, and directly entered the sparse decomposition.

(3) Experimental results showed that SRC-S was better than traditional Chemometrics methods. The possible reason was that the sparse features of spectroscopy helped to distinguish the swill-cooked dirty oil.

(4) SRC-S was slightly better than SRC in the classification which had insufficient and uncertain calibration samples, and it avoided the loss of information during the dictionary training as well.

(5) SRC-S and SRC contained the procedure of sparse decomposition and reconstruction. These procedures had ability to overcome the interference of spectral noise.

Identification between edible oil and swill-cooked dirty oil was one of the most important work in food safety and quality control. Experiment results showed that the NIR technology with SRC-S was an effective method for swill-cooked dirty oil authentication. The SRC-S in this study can also be extended to other applications of NIR technology.

Acknowledgements

This research is financially supported by the National Natural Science Foundation of China (Grant No.51376162), the Scientific Research Projects of Education Department of Zhejiang Provicne (Grant No.201327869), the Public Service Scientific Research Projects of Zhejiang Province (2014C33085), the Natural Science Foundation of Zhejiang Province (No. Q14F050002), and the Inter-disciplinarity Pre-research Project of Zhejiang University of Science and Technology (Grant No.2013JC07Y and No.2014JC07Y).

Notes and references

a College of Information and Electronic Engineering , ZheJiang University of Science and Technology, HangZhou, China. Fax: 0571-85070300; Tel:+8613868077650;E-mail: 109001@zust. edu.cn

b College of Biosystems Engineering and Food Science, ZheJiang University, HangZhou, China;

1. W. Cao, X. Sun, F. Chen, B. XUE, W. CHEN, Q. JIN, and X.WANG, Prospect on discerning technology of swill-cooked dirty oil. China Oils and Fats, 2012, 37: 2.
2. Y.H. Wei, J.Y. Zhang, T.C. Dai, T.H. Tu, and L.G. Luo, Determination of chlorine in hogwash oil and edible oil by ion chromatography. Food.Sci, 2011, 32: 213.
3. R.M. Maggio, L. Cerretani, E. Chiavaro, T.S. Kaufman, and A. Bendini, A novel chemometric strategy for the estimation of extra virgin olive oil adulteration with edible oils. Food Control, 2010, 21:890.
4. E. Lankmayra, J. Mocakb, K. Serdtc, B Ballab, T. Wenzla, D. Bandonienea, M. Gfrerera, and S. Wagnerd, Chemometrical classification of pumpkin seed oils using UV-Vis, NIR and FTIR spectra. Journal of Biochemical and Biophysical Methods, 2004, 61:95.
5. N. Dupuya, O. Galtier, D. Ollivier, P. Vanloota, and J. Artauda, Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis. Analytica Chimica Acta, 2010, 666: 23.
6. M. Casalea, N. Sinelli, P. Oliveri, V. D. Egidiob, and S. Lanteria, Chemometrical strategies for feature selection and data compression applied to NIR and MIR spectra of extra virgin olive oils for cultivar identification. Talanta, 2010, 80: 1832.
7. B. Chen, P. Tian, D.L. Lu, Z. Q. Zhou and M. L. Shao, Feasibility study of discriminating edible vegetable oils by 2D-NIR. Analytical Methods, 2012, 4: 4310.
8. J. Gromadzka, and W. Wardencki. Trends in Edible Vegetable Oils Analysis. Part B. Application of Different Analytical Techniques. Journal of Food and Nutrition Science, 2011, 61:89.
9. J. He, W.T. Xu, Y. Shang, P.Y. Zhu, X.H. Mei, W.Y. Tian, and K.L. Huang, Development and optimization of an efficient method to detect the authenticity of edible oils. Food Control, 2013, 31: 71.
10. M. Bevilacqua, R. Nescatelli, R. Bucci, A. D. Magri, A. L. Magri, and F. Marini, Chemometric Classification Techniques as a Tool for Solving Problems in Analytical Chemistry. Journal of AOAC International, 2014, 97: 19.
11. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y Ma, Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31:210.
12. A. Wagner, J. Wright, A. Ganesh, Z.H. Zhou, H. Mobahi, and Y. Ma, Toward a practical face recognition system: robust alignment and illumination by sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012,34: 372.
13. Jiang. J, Haitao. G, Liangwei. J, Changxin. G, and Nong. S, Semi-supervised Discriminant Analysis and Sparse Representation based self training for Face Recognition. Optik-International Journal for Light and Electron Optics. 2014, 125: 2170.
14. Y. Zhou, S.G. Dai, J. Lv, T.B. Liu, and Y. Shi, Application of Training Dictionary and Sparse Representation in Diesel Property Detection Using Near-infrared Spectroscopy, Acta Chim Sinica. 2012,70:1969.
15. Pingmei C, Guinan W, and Hongjuan Z, K-SVD with reference: an initialization method for dictionary learning. Neural Computing and Applications. 2014, 25:1263.
16. Bo L, Yi S, Zhenghua W, and Jia L, Sufficient conditions for generalized orthogonal matching pursuit in noisy case. Signal Processing. 2015,108:111.
17 Skretting K, and Engan K, Recursive least squares dictionary learning algorithm. IEEE Trans on Signal Processing. 2010,58: 2121.
18. Tropp J.A, Greed is good: Algorithmic results for sparse approximation. IEEE Trans on Information Theory. 2004,50: 2231.
19. Elad M, and Aharon M, Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans on Image Process. 2006, 15: 3736.
20. Marta B, and Federico M, Local classification: Locally weighted-partial least squares-discriminant analysis (LW-PLS-DA). Analytica Chimica Acta. 2014, 838: 20.
21. Paulo H.G, Dias D, Adriano A.G, Marcelo F.P, Beatriz S.F.B, and Mario C.U.A, Simultaneous Classification of Teas According to Their Varieties and Geographical Origins by Using NIR Spectroscopy and SPA-LDA. Food Analytical Methods. 2014,7: 1712.

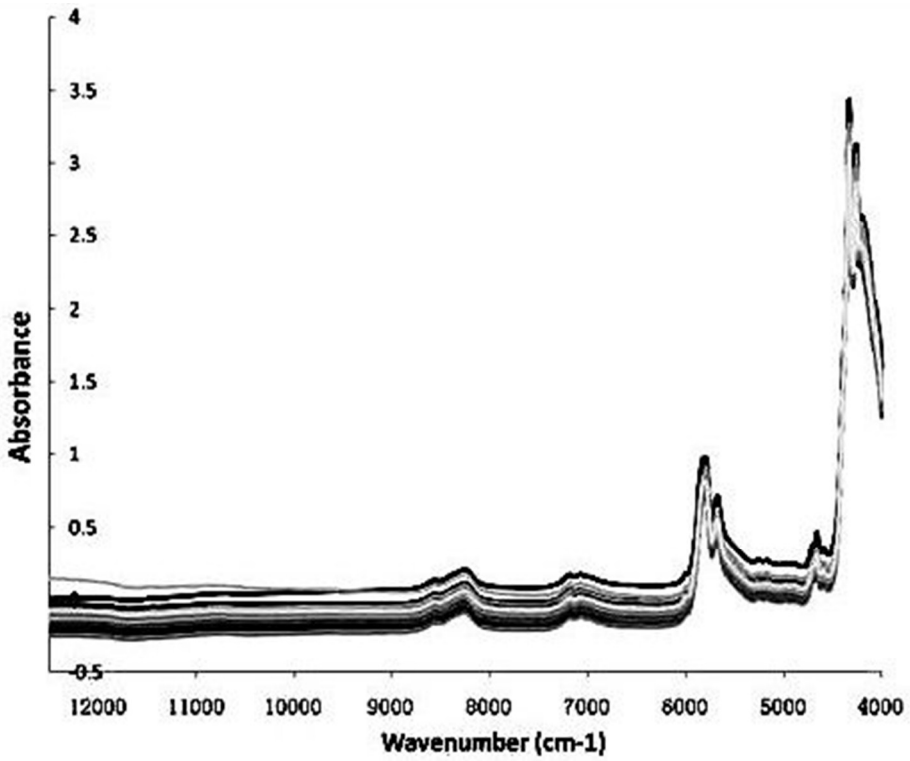


Fig.1. the spectrum of edible oil and swill-cooked dirty oil
160x134mm (72 x 72 DPI)

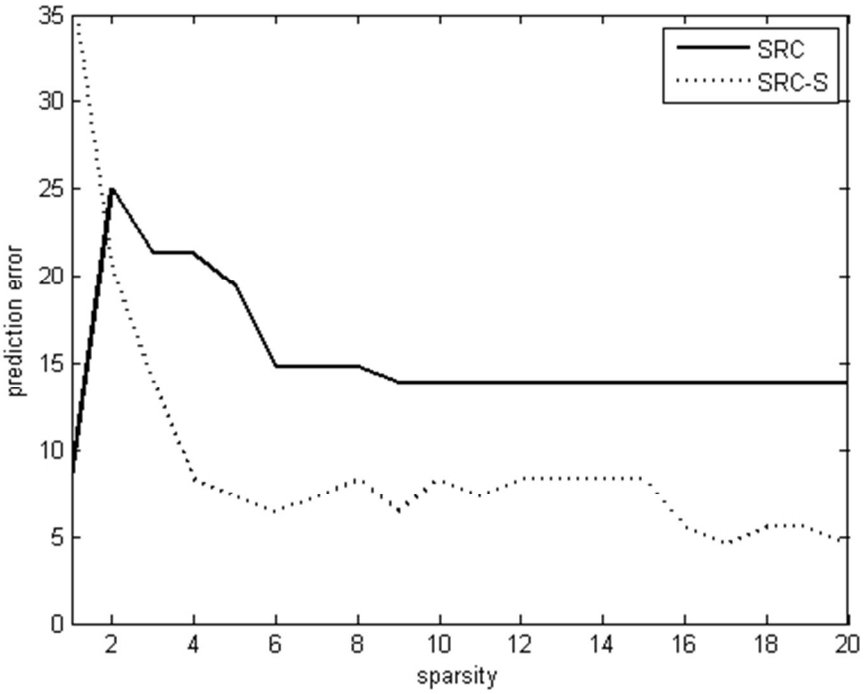


Fig2 the SRC and SRC-S prediction result at different sparsity
148x111mm (96 x 96 DPI)