

基于 FastICA 和神经网络的红酒主要品质参数红外检测

方利民, 林敏*

中国计量学院计量测试工程学院, 浙江 杭州 310018

摘要 为实现红酒中酒精含量、pH 值以及残糖量的快速检测, 对 44 个红酒样品的红外光谱数据进行了分析。使用快速独立分量分析(FastICA)算法对光谱数据矩阵进行分解, 得到独立成分和相应的混合系数矩阵, 再利用误差反向传播算法(back-propagation, BP)构造了三层的神经网络结构, 建立了 ICA-NNR 模型。利用此模型对红酒样品的酒精含量、pH 值以及残糖量进行预测, 根据预测相关系数(r)和预测标准偏差(RMSEP)来评价预测模型的性能, 结果表明该模型对红酒酒精含量、pH 值以及残糖量测定的相关系数 r 分别为 0.953, 0.983 和 0.994, RMSEP 分别为 0.161, 0.017, 0.181。此外, 预测样品集中的 22 个样品 ICA-NNR 模型预测值与参考值相比, 酒精含量、pH 值以及残糖量的最大相对偏差均小于 4%。这为进一步开发红酒成分红外在线分析仪奠定了基础。

关键词 红酒; 快速独立分量分析; 神经网络; 酒精含量; pH 值; 残糖量

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2009)08-2083-04

引言

酒质量的分析检测在酿酒工业的制酒工艺过程中十分重要, 现行使用的一些经典方法, 其中一些是简单的但是精度比较差, 还有一些如测定花色素和丹宁酸是经验性的且费时间的, 而且其结果还需要审视。此外, 酒不是一系列独立的化合物, 而是一些彼此相关化合物的混合物, 例如酒的颜色和酸度及游离 SO_2 有很大的关系。所以在酿酒工业和分析检测中非常需要一种快速、简便的酒质量控制的方法。

近红外光谱分析技术是光谱测量技术与化学计量学的有机结合, 被誉为分析的巨人。它具有快速、可靠、无破坏性、可在线实时检测等特点, 已经成为食品、农业等领域的一种重要检测方法^[1-3]。独立分量分析(independent component analysis, ICA)^[4-6]是近年发展起来的一种全新的数据分析工具, 是解决盲源分离问题的一种有效方法。ICA 方法已经在生物医学信号处理、语音信号处理、图像处理以及光谱数据分析等方面得到了应用^[7-10]。人工神经网络模仿人脑处理信息, 具有分布并行处理、非线性映射、自适应学习和稳健等特性, 对处理灰色、黑色体系与非线性体系的信息有着突出的优点。

本文首先用离散小波变换对 44 个红酒样品的红外光谱

数据进行有效压缩, 然后用 FastICA 算法对其进行分解, 提取独立成分和相应的混合系数矩阵, 最后用 BP 神经网络建立红酒中酒精含量、pH 值以及残糖量三个主要品质参数的检测模型。结果表明, 校正样品集的红外模型预测值与参考值的相关系数都达到 0.95 以上, 所建模型具有实用价值。通过比较, 用 ICA-NNR 所建模型在相关系数与均方根误差两个指标上均优于 PCR 和 PCA-NNR 方法。

1 FastICA 算法

ICA 理论及分离算法的关键在于如何度量分离结果的独立性, 本文将应用一种基于负熵的独立性判决准则的 FastICA 算法对光谱数据进行处理, 该算法采用牛顿迭代算法, 收敛速度快。它依据负熵判据来寻找分离矩阵 W 以实现独立分量的提取。使

$$J(y) \propto [E\{G(y)\} - E\{G(y_{\text{gauss}})\}]^2 \quad (1)$$

最大, 其中 $J(y)$ 为 y 的负熵; $G(\cdot)$ 为一种非线性、非二次的函数^[6]。由牛顿迭代可得如下定点 ICA 算法

$$\begin{cases} w^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w \\ w^* = w^+ / \|w^+\| \end{cases} \quad (2)$$

式中 $g(\cdot)$ 是 $G(\cdot)$ 导数, w^+ 为某一次牛顿迭代的结果; w^* 为 w^+ 归一化后的值。每次提取一个分量之后, 需要从中减

收稿日期: 2008-09-08, 修订日期: 2008-12-10

基金项目: 国家自然科学基金项目(50675214)和浙江省科技厅重点项目(2006C21044)资助

作者简介: 方利民, 1983 年生, 中国计量学院计量测试工程学院硕士研究生 e-mail: fanglm1004@163.com

* 通讯联系人 e-mail: linm@cjl.u.edu.cn

去该独立分量,直到所有分量都被提取出来,方法如下

$$\begin{cases} \omega_{k+1} = \omega_{k+1} - \sum_{j=1}^k \omega_{k+1}^T \omega_j \omega_j \\ \omega_{k+1} = \omega_{k+1} / \sqrt{\omega_{k+1}^T \omega_{k+1}} \end{cases} \quad (3)$$

式中 ω_{k+1} 是第 $k+1$ 次牛顿迭代的结果(假设已经估计 k 个分量)。

2 光谱模型及求解算法

根据 Beer-Lambert 定律,对于 IR 光谱数据矩阵 A ,可建模为各成分的光谱信号与其贡献度乘积的加和

$$A = MI \quad (4)$$

其中, I 是独立成分矩阵,在理想的分解状态下相当于纯物质的光谱矩阵, M 是混合矩阵,它与纯物质在混合样品中的浓度有关。本文将 ICA 算法与神经网络相结合,用于红酒的红外光谱解析,如图 1 所示。

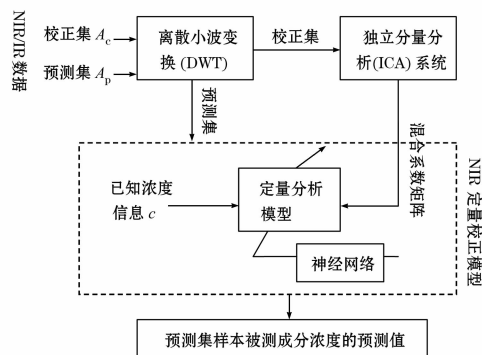


Fig. 1 ICA-NNR scheme

3 试验数据及结果

3.1 试验数据

数据为 University of Copenhagen 提供的红酒样品红外光谱数据以及相应的酒精含量(单位: vol%)、pH 值以及残糖量(单位: $\text{g} \cdot \text{L}^{-1}$)^[11]。总共有来自不同地区、同一种类葡萄所酿制的 44 个红酒样品,经福斯公司的葡萄酒扫描分析器(WineScan™ FT120 Basic, FOSS Analytical)在波数 5 011 ~ 929 cm^{-1} 范围内扫描得到红外光谱,采样间隔为 4 cm^{-1} ,每个样本光谱含有 1 056 个数据点。在建模之前将样品划分为校正样品集和预测样品集,各取总样品数的 1/2。图 2 为某一红酒样品的 IR 谱图。在 1 545 ~ 1 710 cm^{-1} (a) 和 2 968 ~ 3 620 cm^{-1} (b) 波段出现了较大的噪声,该噪声的产生与水的吸收带有关。

在预测过程中,均以方根误差(RMSEP)和相关系数(r)的大小作为模型预测准确度的评价。RMSEP 的数值越小, r 越大,模型的预测准确度越高。其中 RMSEP 的定义如下

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\tilde{y}_i - y_i)^2}{n}} \quad (5)$$

其中 \tilde{y}_i 为 y_i 的预测值。

3.2 神经网络结构

采用 3 层 BP 神经网络,传递函数分别为 Tansig 和

Purelin 函数,优化学习算法选用 Levevberg-Marquardt 算法。对酒精含量、pH 值以及残糖量分别建立校正模型。

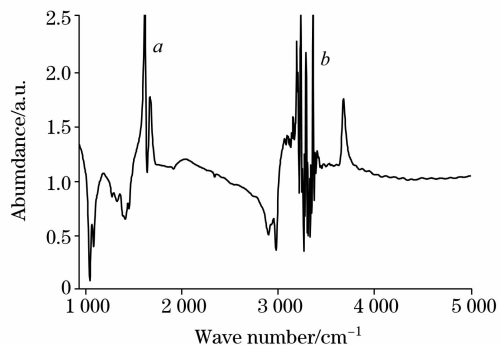


Fig. 2 Typical FTIR spectrum of a red wine sample

首先,通过分析被测含量校正模型的 RMSEP 值与隐层神经元 Nods 之间的关系,确定了酒精含量、pH 值以及残糖量的定量分析模型中 Nods 的取值分别为 Nods=6, 7, 7。其次,网络的输入为经 FastICA 算法分解得到的混合系数矩

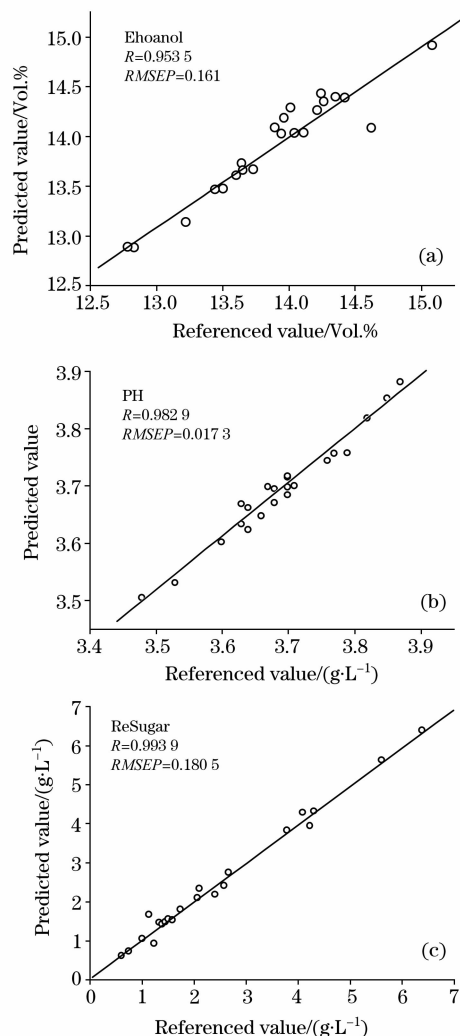


Fig. 3 Correlations between predicted value and referenced value of ethanol (a), pH (b) and rest sugar (c) contents in red wine samples

阵, 因此 ICA 的分量数 ICs 不仅关系到算法的精确度, 而且与网络的运行速度以及所建模型的精度有关。本文采用不同的分量数选取来得到最优的数值。分析三种被测含量校正模型的性能和计算量, 确定了酒精含量、pH 值以及残糖量的定量分析模型中的 ICs 值分别为 ICs=4, 5, 4。网络的输出层节点数取为 1。

3.3 试验结果

本文采用 ICA-NNR 方法, 选取上述网络结构, 分别建立了红酒中的酒精含量、pH 值以及残糖量的定量分析模型。校正集样品的预测值与所提供参考值之间的相关性分析见

图 3。可以看出, 校正模型的相关系数 r 都在 0.95 以上, pH 值和残糖量的测定可达到 0.98 和 0.99, 校正模型的均方根误差也都比较小, 这都充分地说明了该模型的线性特征很好。

下面用建立起来的定量模型对预测集中未知样品进行了预测分析。用 IR 结合 ICA-NNR 方法所得到的数据与所提供参考值吻合很好, 如表 1 所示。可见, 预测样品集中的 22 个样品 IR 方法测定值与参考值相比, 酒精含量、pH 值以及残糖量的最大相对偏差均小于 4%。由于 BP 神经网络的自组织、自学习性对样品中随机误差的校正, 预测精度得到了显著提高, 此外模型的稳定性也会有所增强。

Table 1 Comparison between predicted value and referenced value of unknown samples

No.	Ethanol/vol%			pH			Rest Sugar/(g · L ⁻¹)		
	True	IR	δ/%	True	IR	δ/%	True	IR	δ/%
1	13.74	13.65	0.60	3.87	3.88	0.26	0.62	0.61	1.18
2	13.65	13.72	0.51	3.60	3.60	0.02	1.06	1.05	0.19
3	15.09	14.90	1.24	3.67	3.69	0.73	1.15	1.15	0.12
4	14.43	14.37	0.36	3.70	3.69	0.10	1.16	1.17	0.58
5	13.97	14.17	1.47	3.64	3.66	0.55	1.37	1.40	2.36
6	13.66	13.65	0.08	3.68	3.69	0.36	1.45	1.44	0.18
7	13.23	13.12	0.77	3.77	3.75	0.40	1.52	1.55	2.46
8	13.51	13.46	0.33	3.63	3.63	0.05	1.60	1.61	0.64
9	14.05	14.02	0.20	3.82	3.81	0.09	1.75	1.71	2.29
10	13.90	14.07	1.28	3.66	3.64	0.38	2.08	2.07	0.31
11	14.36	14.38	0.19	3.85	3.85	0.04	2.19	2.21	0.93
12	13.95	14.01	0.48	3.79	3.75	0.90	2.36	2.34	0.56
13	14.12	14.02	0.65	3.70	3.71	0.36	2.45	2.46	0.51
14	14.63	14.07	3.80	3.53	3.53	0.02	2.59	2.59	0.21
15	13.45	13.45	0.06	3.48	3.50	0.67	2.68	2.65	0.80
16	12.84	12.87	0.25	3.71	3.70	0.30	2.74	2.72	0.62
17	14.27	14.34	0.49	3.76	3.74	0.47	3.80	3.78	0.38
18	13.61	13.60	0.08	3.64	3.62	0.50	4.10	4.11	0.22
19	12.79	12.88	0.70	3.70	3.68	0.47	4.24	4.17	1.50
20	14.02	14.28	1.83	3.68	3.67	0.30	4.32	4.34	0.58
21	14.22	14.25	0.22	3.63	3.66	1.02	5.62	5.63	0.12
22	14.25	14.42	1.20	3.70	3.71	0.42	6.40	6.38	0.23
δ _{max}			3.80			1.02			2.46

δ/%: relative deviation; δ_{max}: maximum relative deviation

Table 2 Comparison of the calibration models of main quality parameters of red wine constructed by PCR, PCA-NNR and ICA-NNR methods

Items	Ethanol/vol%			pH			Rest Sugar/(g · L ⁻¹)		
	PCR	PCA-NNR	ICA-NNR	PCR	PCA-NNR	ICA-NNR	PCR	PCA-NNR	ICA-NNR
Parameters	PCs=15	PCs=11 Nods=6	ICs=4 Nods=6	PCs=17	PCs=9 Nods=7	ICs=5 Nods=7	PCs=10	PCs=9 Nods=7	ICs=4 Nods=7
RMSEP	0.316	0.306	1.5	0.025	0.038	0.017	0.694	0.506	0.181
r	0.818	0.910	0.989 4	0.963	0.914	0.983	0.871	0.921	0.994

PCs: PCA(The number of principal components in PCA)

下面用现行常用的 PCR 方法, 以及 PCA 分解结合神经网络回归(PCA-NNR)方法分别建立红酒主要品质参数的校正模型, 并与本文所用方法作比较。模型主要参数和评价指

标的比较结果见表 2。其中 PCA-NNR 方法所用的神经网络结构与 ICA-NNR 方法是一样的。比较 PCR 与 PCA-ANN 方法, 前者是线性的, 后者则是非线性的。对酒精含量和残糖

量的测定, PCA-ANN 要优于 PCR, 而对于 pH 值的测定则是后一方法更好, 说明 pH 值与红酒光谱之间关系更偏向线性。此外, ICA-NNR 方法所建模型在 RMSEP 和 r 两个指标上均优于 PCR 和 PCA-NNR 方法。

4 结 论

应用红外光谱技术对红酒中酒精含量、pH 值以及残糖量进行了预测, 采用 BP 神经网络与 FastICA 算法相结合的

方法, 选择网络最佳参数, 建立了 ICA-NNR 模型。预测结果的相关系数和均方根误差都达到了较好的效果, 是对传统红酒指标测定方法的改进, 为开发出更高精度的红酒酒精含量、pH 值、残糖量以及其他一些指标的检测仪器以及实时在线测量提供了依据。将 ICA 方法与 IR 技术应用于红酒品质的检测, 丰富了化学计量学方法, 也拓宽了 ICA 算法的应用领域。对于今后的研究, 主要是优化预测模型, 提高模型预测能力和稳健性。

参 考 文 献

- [1] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). Foundation of Near Infrared Spectroscopy and Its Application(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005. 1.
- [2] YU Hai-yan, YING Yi-bin, FU Xia-ping, et al(于海燕, 应义斌, 傅霞萍, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(5): 920.
- [3] SHAO Yong-ni, HE Yong(邵咏妮, 何 勇). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2006, 25(6): 478.
- [4] Hyvärinen A, Oja E. Neural Computation and Applications, 1997, 9(7): 1483.
- [5] Hyvärinen A, Oja E. Neural Networks A, 2000, 13(4-5): 411.
- [6] Hyvärinen A. IEEE Transactions on Neural Networks, 1999, 10(3): 626.
- [7] BI Xian, LI Tong-hua, WU Liang(毕 贤, 李通化, 吴 亮). Chemical Journal of Chinese Universities(高等学校化学学报), 2004, 25(6): 1023.
- [8] Chen J, Wang X Z. Journal of Chemical Information and Computer Sciences, 2001, 41: 992.
- [9] FANG Li-min, LIN Min(方利民, 林 敏). Chinese Journal of Analytical Chemistry(分析化学), 2008, 36(6): 815.
- [10] Hahn S, Yoon G. Applied Optics, 2006, 45: 8374.
- [11] Thomas S, Davide B, Rasmus B. Analytical Chimica Acta, 2008, 615: 18.

Detection of the Main Quality Indicators in Red Wine with Infrared Spectroscopy Based on FastICA and Neural Network

FANG Li-min, LIN Min*

College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China

Abstract For the rapid detection of the ethanol, pH and rest sugar in red wine, infrared (IR) spectra of 44 wine samples were analyzed. The algorithm of fast independent component analysis (FastICA) was used to decompose the data of IR spectra, and their independent components and the mixing matrix were obtained. Then, the ICA-NNR calibration model with three-level artificial neural network (ANN) structure was built by using back-propagation (BP) algorithm. The models were used to estimate the contents of ethanol, pH and rest sugar in red wine samples for both in calibration set and predicted set. Correlation coefficient (r) of prediction and root mean square error of prediction (RMSEP) were used as the evaluation indexes. The results indicate that the r and RMSEP for the prediction of ethanol content, pH and rest sugar content are 0.953, 0.983 and 0.994, and 0.161, 0.017 and 0.181, respectively. The maximum relative deviations between the ICA-NNR method predicted value and referenced value of the 22 samples in predicted set are less than 4%. The results of this paper provide a foundation for the application and further development of IR on-line red wine analyzer.

Keywords Red wine; Fast ICA; Neural network; Content of ethanol; pH; Rest sugar

(Received Sep. 8, 2008; accepted Dec. 10, 2008)

* Corresponding author