

腐乳感官和理化品质的核主成分分析

经玲¹ 朱甫芹¹ 鲁绯² 孙君社²

(1. 中国农业大学理学院,北京 100083; 2. 中国农业大学食品科学与营养工程学院,北京 100083)

摘要 为解决 PCA 不适合多指标综合分析中非线性主成分分析的问题,采用核主成分分析(KPCA)方法,对我国不同地区 16 种腐乳的品质进行了综合评价。使用核函数将原空间映射到高维特征空间,在高维空间进行了线性主成分分析。结果表明,通过对核参数的适当选取,可使最大特征值的贡献率达到或接近 85%,避免了多个主成分的不同组合而导致的评价结果的不一致。应用 KPCA 得到的腐乳品质的评价结果具有一定的客观性。

关键词 核主成分分析;腐乳品质;综合评价

中图分类号 TS 214.2; TP 391

文章编号 1007-4333(2004)03-0079-03

文献标识码 A

Kernel principal composition analysis of sensory and physiochemical quality of fermented bean curd

Jing Ling¹, Zhu Fuqin¹, Lu Fei², Sun Junshe²

(1. College of Science, China Agricultural University, Beijing 100083, China;

2. College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China)

Abstract Through kernel principal composition analysis, a new method of comprehensive evaluation was made for the sensory and physiochemical quality of sixteen kinds of fermented bean curd from markets. By using the kernel functions, one can efficiently calculate principal compositions in high dimensional feature spaces, related in input space by some nonlinear map. The results showed that the maximum eigenvalue contributes nearly 85% by choosing appropriate parameters, avoiding the different array as a result of many principal composition. KPCA should have a good performing in comprehensive analysis of the quality of fermented bean curd.

Key words kernel principal composition analysis (KPCA); sensory and physiochemical quality; comprehensive evaluation

腐乳是我国传统大豆发酵制品。20 世纪 70 年代以来,国内外有关腐乳生产菌种选育、酿造过程的生化反应以及工艺改进等方面的研究较多,但在腐乳感官品质描述及综合评价方面的研究较少。杨坚^[1]以各地生产的有代表性的 16 种白腐乳为试验材料,采用多指标描述了腐乳的感官和理化品质,并进行了主成分分析和综合评价。主成分分析法(PCA)是将多个变量综合成少数变量的一种多元统计方法。PCA 作为一种线性的降维方法,为解决多指标的综合评价提供了很好的手段,但是在现实情况下,指标之间的关系往往是非线性的。另外,当指

标之间的相关性较小时,可能出现各指标贡献率过于分散的情况,此时若取较多的主成分,不同的组合将会导致评价结果的不一致,从而影响评价效果。因此,要综合相关性不大的变量,采用线性主成分分析法显然是不妥的。为此研究者提出了一系列的改进 PCA,如文献[2]提出的最小窗口 PCA(MWPCA),由于其本质上是用逐段线性去逼近非线性,因此无法从根本上解决非线性问题。自 20 世纪 90 年代,尤其是支持向量机(SVM)的研究展开后,关于核方法的研究受到重视。最新研究成果^[3~5]表明,核主成分分析法(KPCA)通过 PCA 与核方法的有

收稿日期:2003-10-31

基金项目:国家自然科学基金资助项目(10371131)

作者简介:经玲,副教授,博士,主要从事运筹与优化、数据挖掘等方面的研究,E-mail:jingling_aaa@163.com

机融合,可以将观察变量空间 X 通过一个非线性变换 映射到高维的特征空间 F ,在 F 内进行线性主成分分析。通过核的技巧, KPCA 评价方法只需在原空间进行点积计算,而不必知道 的确切形式,巧妙地解决了非线性 PCA 中非线性变换的未确知性,并且避免了多个主成分的不同组合而导致的评价结果的不一致,因此特别适合处理非线性问题,且能提供更多的信息。目前 KPCA 已较好地应用于科技期刊质量、家庭消费状况等的综合评价^[6,7],但在食品品质综合评价方面还未见应用。本文中提出的腐乳品质的 KPCA 评价方法具有通用性,其目的不在于对市售的 16 种腐乳的品质进行排名,旨在从方法论角度进行一些尝试性探讨,以提高腐乳品质评价的科学性、有效性和客观性。

1 KPCA 的基本原理

设 $x_i \in \mathbf{R}^N (i = 1, 2, \dots, n)$ 为样本数据向量,把输入空间 \mathbf{R}^N 通过非线性变换 投影到特征空间 F ,即

$$\mathbf{R}^N \rightarrow F \quad (1)$$

假定特征空间 $(x_1), (x_2), \dots, (x_n)$ 满足

$\sum_{i=1}^n (x_i) = 0$,非线性 PCA 就是在 F 中对下面的协方差矩阵进行线性 PCA :

$$C = \frac{1}{n} \sum_{j=1}^n (x_j) (x_j)^T \quad (2)$$

现求 C 的特征值 (λ_i) 和特征向量 $F \setminus \{0\}$ 。

$$C \cdot v = \lambda v \quad (3)$$

即有 $(v_k) \cdot C = \lambda_k (v_k)$, $k = 1, 2, \dots, n$ 。

由于式 (3) 的所有解均在 $(x_1), (x_2), \dots, (x_n)$ 张成的子空间内,因此,存在系数 $\alpha_1, \alpha_2, \dots,$

$$v_k = \sum_{i=1}^n \alpha_i (x_i) \quad (4)$$

其中

$$K = (k_{ij})_{n \times n} = ((x_i) \cdot (x_j)) \quad i, j = 1, 2, \dots, n \quad (5)$$

求解式 (4) 得到特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$, 以及对应的特征向量 v_1, v_2, \dots, v_n 。由于

$$\begin{aligned} (v_k \cdot v_k) &= \sum_{i,j=1}^n \alpha_i \alpha_j ((x_i) \cdot (x_j)) = \\ (v_k \cdot K v_k) &= \lambda_k (v_k \cdot v_k) \end{aligned} \quad (6)$$

其中 $\alpha_i = \frac{1}{\sqrt{\lambda_k}} ((x_i) \cdot v_k)$, 因此可以通过标准化 v_k 来

使 $(v_k \cdot v_k) = 1$ 。

对于主成分的选取,只需计算一个测试点 (x) 在 F 的特征向量 v_k 上的投影

$$(v_k \cdot (x)) = \sum_{i=1}^n \alpha_i ((x_i) \cdot (x)) = \sum_{i=1}^n \alpha_i K(x_i, x) \quad (7)$$

式 (7) 形式上类似于 PCA 中的综合指标,也称式 (7) 为 KPCA 的第 k 个主成分。此时 KPCA 的综合评价函数是

$$F(x) = \sum_{k=1}^r \sum_{i=1}^n \alpha_i v_k K(x_i, x) \quad (8)$$

其中 r 满足 $\sum_{i=1}^r \lambda_i / \sum_{j=1}^n \lambda_j \geq 80\%$, v_k 为对应的第 k 个主成分的贡献率。通常,若核函数中的参数选择合适,能够做到 $r = 1$ 。这样式 (8) 就简化为

$$F(x) = \sum_{i=1}^n \alpha_i K(x_i, x) \quad (9)$$

若 $\sum_{i=1}^n (x_i) = 0$, 则式 (4) 中的 K 可以用 K^* 表示:

$$K^* = K - \mathbf{1}_n \mathbf{1}_n^T - \mathbf{K} \mathbf{1}_n + \mathbf{1}_n \mathbf{K} \mathbf{1}_n \quad (10)$$

其中 $(\mathbf{1}_n)_{ij} = 1/n$ 。

2 腐乳感官和理化品质的综合评价

本文中 16 种腐乳的感官评定指标和理化特性指标数据均引自文献[1],其中的感官评定结果采用专家和评审员评定结果相结合的方法得出。

2.1 腐乳感官品质的核主成分分析

选取多项式核为本试验的核函数,即 $K(x, y) = [c(x \cdot y) + m]^d$; 取 $c = 0.0016, m = 0.01, d = 30$, 计算 $K^*/16$ 的特征值(表 1)及相应的感官样本得分(表 2)。

表 1 腐乳感官品质的 $K^*/16$ 的特征值及贡献率

Table 1 The eigenvalues of $K^*/16$ and cumulative percentage for sensory attributes of fermented bean curd

特征值序号	特征值	贡献率/ %	累积贡献率/ %
1	3.9164	84.82	84.82
2	0.4380	9.49	94.31
3	0.1140	2.47	96.78
4	0.0519	1.12	97.90
5	0.0363	0.79	98.69
6	0.0260	0.56	100
...

2.2 腐乳理化品质的核主成分分析

仍选取多项式核为本试验的核函数, 取 $c = 0.0002$, $m = 0.002$, $d = 20$, 采用同样的方法计算此处 $K^*/16$ 的特征值, 得最大特征值 $= 0.6917$, 贡献率 86.31% 及相应的样本得分。感官和理化品质的综合得分 (取算术平均) 见表 2。

表 2 16 种腐乳样本的感官品质、理化品质和综合得分

Table 2 Sensory scores, physicochemical scores and comprehensive scores of the quality of 16 fermented bean curd

样本 编号	感官品质		理化品质		综合得分	
	得分	名次	得分	名次	得分	名次
1	0.7032	1	-0.1499	9	0.2767	9
2	0.2842	12	0.2863	5	0.2858	8
3	0.5431	7	0.2809	6	0.4120	5
4	0.6473	4	1.1814	2	0.9144	2
5	0.6387	5	0.2913	4	0.4650	3
6	0.4810	8	1.6734	1	1.0772	1
7	0.5698	6	0.3279	3	0.4489	4
8	0.6541	3	0.1026	7	0.3784	6
9	-6.2038	16	-0.8986	14	-3.5512	16
10	-0.0888	14	-0.2127	11	-0.1507	13
11	0.6649	2	0.0312	8	0.3481	7
12	0.2828	13	-0.3657	13	-0.0414	12
13	-0.3339	15	-0.9560	15	-0.6450	15
14	0.3418	10	-0.3538	12	-0.0060	11
15	0.4786	9	-1.0394	16	-0.2804	14
16	0.3369	11	-0.1990	10	0.0689	10

3 结 论

利用核的技巧, KPCA 评价方法可以在特征空

间实施线性 PCA, 只需在原输入空间进行点积计算即可, 巧妙地解决了非线性 PCA 中非线性变换的不确定性。利用 KPCA 评价方法对 16 种市售腐乳感官和理化品质的综合评价结果表明, 通过对核参数的适当选取, 可使最大特征值的贡献率达到或接近 85% , 避免了 PCA 多个主成分的不同组合而导致的评价结果的不一致。另外, KPCA 进行的是非线性主成分分析, 因而评价结果较为客观。

参 考 文 献

- [1] 杨坚, 童华荣, 贾利蓉. 豆腐乳感官和理化品质的主成分分析[J]. 农业工程学报, 2002, 18(2): 131~135
- [2] 赵立杰, 王纲, 李元. 非线性主成分分析故障检测和诊断方法及应用[J]. 信息与控制, 2001, 30(4): 359~364
- [3] Scholkorf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(11): 1299~1319
- [4] Muller K R, Mika S, Ratsch S, et al. An introduction to kernel-based learning algorithms[J]. IEEE Trans on Neural Network, 2001, 12(2): 181
- [5] Twining C, Taylor C. The use of kernel principal component analysis to model data distributions [J]. Pattern Recognition, 2003, 36(1): 217~227
- [6] 张爱丽, 刘广利, 刘清水. 科技期刊综合评价模型- KPCA[J]. 计算机工程与应用, 2003(24): 200~201
- [7] 肖健华, 吴今培. 基于核的特征提取技术及应用研究[J]. 计算机工程, 2002, 28(10): 36~38