

撤稿声明

撤稿文章名: 多维标度法在可视化卷烟感官评价结果中的应用
作者: 申小征, 卫慧芳

* 通讯作者: 邮箱: 1528435114@qq.com

期刊名: 统计学与应用 (SA)
年份: 2016
卷数: 5
期数: 1
页码 (从X页到X页): 28-38
DOI (to PDF): <http://dx.doi.org/10.12677/sa.2016.51004>
文章ID: 2580156
文章页面: <https://www.hanspub.org/journal/PaperInformation.aspx?paperID=17250>
撤稿日期: 2018-6-25

撤稿原因 (可多选):

- 所有作者
 部分作者:
 编辑收到通知来自于
- 出版商
 科研机构:
 读者:
 其他:

撤稿生效日期: 2018-6-25

撤稿类型 (可多选):

- 结果不实
 实验错误 数据不一致 分析错误 内容有失偏颇
 其他:
 结果不可再得
 未揭示可能会影响理解与结论的主要利益冲突
 不符合道德
- 欺诈
 编造数据 虚假出版 其他:
 抄袭 自我抄袭 重复抄袭 重复发表 *
 侵权 其他法律相关:
- 编辑错误
 操作错误 无效评审 决策错误 其他:
- 其他原因:

出版结果 (只可单选)

- 仍然有效.
 完全无效.

作者行为 失误(只可单选):

- 诚信问题
 学术不端
 无 (不适用此条, 如编辑错误)

* 重复发表: "出版或试图出版同一篇文章于不同期刊."

历史

作者回应:

是, 日期: yyyy-mm-dd

否

信息改正:

是, 日期: yyyy-mm-dd

否

说明:

“多维标度法在可视化卷烟感官评价结果中的应用”一文刊登在2016年3月出版的《统计学与应用》2016年第5卷第1期第28-38页上。因作者个人失误, 文章部分内容有待进一步确认, 根据国际出版流程, 编委会现决定撤除此重复稿件, 保留原出版出处:

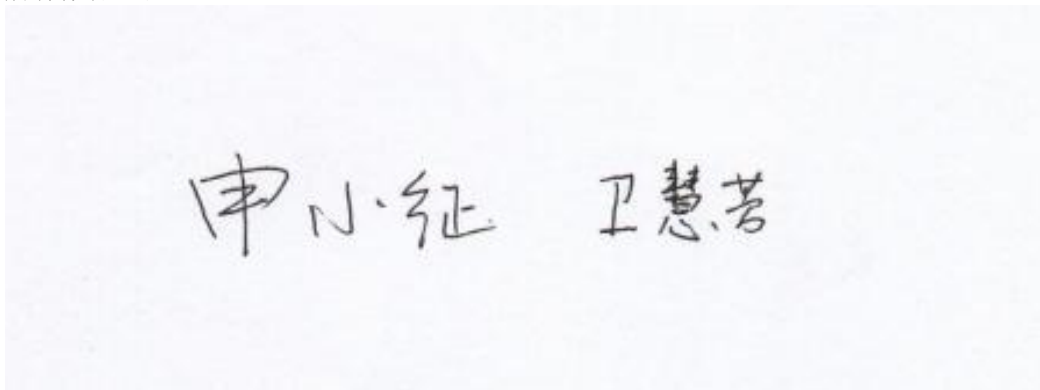
申小征, 卫慧芳. 多维标度法在可视化卷烟感官评价结果中的应用[J]. 统计学与应用, 2016, 5(1): 28-38.

<http://dx.doi.org/10.12677/SA.2016.51004>

指导编委:

Firstname Lastname
(function e.g. EiC, journal abbreviation)

所有作者签名:



申小征 卫慧芳

Application of Multidimensional Scaling in Visual Cigarette Sensory Evaluation Results

Xiaozheng Shen, Huifang Wei

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: 1528435114@qq.com

Received: Mar. 8th, 2016; accepted: Mar. 27th, 2016; published: Mar. 30th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Cigarette sensory evaluation is a multi index evaluation, so that the various aspects of cigarette sensory can be characterized comprehensively. However, the evaluation results are multi-dimensional, and it is not easy to compare the differences of different cigarettes directly. For the national evaluation results of 54 representative samples of cigarettes, there is a comparative study of various multivariate statistical mapping techniques. The results showed that: 1) Some lossless graphing methods such as Facebook statistics chart, star chart, parallel coordinate plots [1] [2] so on are only suitable for visual comparison in a small number of samples; 2) In the nondestructive mapping method, when using principal component analysis, the variance proportion of the first few principal components explained is relatively small. It means that we need at least ten dimensions space, in order to keep the original data as more information as possible. Therefore, the use of principal component analysis to visualize severely distorted in two-dimensional space; using multidimensional scaling method [3] will be able to better preserve the original space dissimilarity between two samples in four-dimensional space. In conclusion, multidimensional scaling method is more suitable for visualization of cigarette sensory evaluation results.

Keywords

Multidimensional Scaling Method, Principal Component Analysis, Sammon Nonlinear Mapping, Nondestructive Mapping

多维标度法在可视化卷烟感官评价结果中的应用

申小征, 卫慧芳

云南财经大学统计与数学学院, 云南 昆明
Email: 1528435114@qq.com

收稿日期: 2016年3月8日; 录用日期: 2016年3月27日; 发布日期: 2016年3月30日

摘要

卷烟感官评价是一种多指标评价, 以便较全面表征卷烟的各方面感官特征。但由于评价结果为多维, 不易于直接比较多种不同卷烟之间的不相似度。为此, 对全国54个代表性卷烟样品的评价结果, 比较研究了多种多元统计作图技术的表现。结果表明: 1) 统计脸谱图、星形图、平行坐标图[1][2]等无损作图法仅适合少数样品的可视化比较; 2) 无损作图法中, 采用主成分分析法时, 前几个主成分解释的方差比例比较小, 需要至少在十维空间才能保持原始数据的信息不会有严重损失。因此, 采用主成分分析在二维空间中进行可视化会严重失真; 采用多维标度法[3]在四维空间就能较好保留原始空间中两两样品的不相似度, 因此, 多维标度法更适合对卷烟感官评价结果进行可视化。

关键词

多维标度法, 主成分分析, Sammon非线性映射, 无损作图法

1. 引言

卷烟感官分析评价技术是在理化分析的基础上, 集心理学、生理学、统计学和专业知识为一体, 通过长期的实践而发展起来的一门学科。卷烟感官分析与评价是通过评价员的视觉、嗅觉、味觉等感受, 而引起的反应并感知到卷烟或卷烟中其它物质的特征或性质的一种科学方法。因此卷烟感官评价是一种多指标评价, 以便较全面表征卷烟的各方面感官特征。基于应用卷烟风格感官评价方法, 以卷烟的风格特色为切入点, 对中式卷烟进行品类细分与集合, 不失为一种中式卷烟品类构建的系统方法。

可视化评吸数据可以直观的以图形表现形式观察哪些样品相近, 哪些隔得较远, 从而达到对中式卷烟进行品类细分的目的。另一方面, 可视化评吸数据可以方便地用来对聚类方法的聚类结果进行初步的直观评估, 观察相近的样品是不是聚到一起。但由于评价结果为多维, 不易于直接比较多种不同卷烟之间的不相似度, 因此我们需要多元统计作图技术来对样品进行可视化。

多元统计提供了多种作图技术, 1) 无损作图法: 统计脸谱法、星形图、平行坐标图等; 2) 有损作图法: 当卷烟样品处于高维空间中, 直接可视化是不现实的, 一个自然地想法是在不严重损失原始数据信息的条件下, 将数据投影到低维空间(2、3 维空间)进行可视化, 主成分分析是一个常用的线性数据维数约简方法; 另一种是非线性维数约简方法: 多维标度法; 本文对这些进行了比较研究, 旨在寻求一种适用于可视化多维感官评价结果的作图技术。

2. 样品与方法

2.1. 样品

中式卷烟风格感官评价方法研究的主体对象是中式卷烟代表性品牌产品, 各卷烟企业不同品牌、不同风格特征的产品, 完整地组成了中式卷烟大家族。项目组对样品清单进行了相关的数据指标分析, 提出了入选样品的原则和要求并进行了初选。具体要求为: a) 所选品牌为重点骨干品牌; b) 所选品牌规格能够代表中式卷烟风格特色; c) 卷烟样品存放在1年之内; d) 样品数量不宜太多。最终确定了54个中式卷烟代表产品(见表1), 作为方法研究相对固定的主体对象。其中, 烤烟型卷烟样品50个、混

Table 1. Sample list of sensory evaluation of Chinese style cigarettes

表 1. 中式卷烟风格感官评价样品清单

序号	生产企业	卷烟规格	条形码	焦油量(mg)
1	广东中烟工业公司	双喜(软)	6901028001687	13
2	广东中烟工业公司	双喜(硬经典)	6901028000659	13
3	广东中烟工业公司	五叶神(硬金)	6901028007238	13
4	福建中烟工业公司	七匹狼(豪情)	6901028138369	13
5	福建中烟工业公司	七匹狼(红)	6901028137492	13
6	福建中烟工业公司	金桥(软混)	6901028143080	12
7	河南中烟工业公司	黄金叶(世纪之光)	6901028163538	13
8	河南中烟工业公司	红旗渠(硬金红)	6901028164610	13
9	河南中烟工业公司	帝豪(硬金黄)	6901028170758	13
10	江西中烟工业公司	金圣(硬)	6901028133302	13
11	江西中烟工业公司	金圣(吉品)	6901028133890	13
12	内蒙古昆明卷烟有限责任公司	云烟(12mg 芙蓉)	6901028177122	12
13	甘肃烟草工业有限责任公司	兰州(硬珍品)	6901028065504	11
14	吉林烟草工业有限责任公司	长白山(红)	6901028100366	13
15	吉林烟草工业有限责任公司	长白山(神韵)	6901028099783	5
16	深圳烟草工业有限责任公司	好日子(软珍品)	6901028942461	13
17	上海烟草(集团)公司	中华(硬)	6901028075015	13
18	上海烟草(集团)公司	中华(软)	6901028075022	13
19	上海烟草(集团)公司	红双喜(硬)	6901028075084	13
20	上海烟草(集团)公司	中南海(10mg)	6901028071291	10
21	湖南中烟工业有限责任公司	芙蓉王(硬)	6901028193504	13
22	湖南中烟工业有限责任公司	芙蓉王(蓝)	6901028193863	12
23	湖南中烟工业有限责任公司	白沙(硬)	6901028191142	13
24	湖南中烟工业有限责任公司	白沙(精品 2 代)	6901028191111	13
25	贵州中烟工业有限责任公司	贵烟(多彩)	6901028037938	12
26	贵州中烟工业有限责任公司	黄果树(佳品)	6901028036962	13
27	红云红河烟草(集团)有限责任公司	云烟(软珍品)	6901028045919	13
28	红云红河烟草(集团)有限责任公司	云烟(紫)	6901028046893	13
29	红云红河烟草(集团)有限责任公司	红河(硬甲)	6901028055345	13
30	黑龙江烟草工业有限责任公司	林海灵芝(软白)	6901028103183	13
31	湖北中烟工业有限责任公司	黄鹤楼(软蓝)	6901028180580	12
32	湖北中烟工业有限责任公司	黄鹤楼(硬雅香)	6901028179867	12
33	湖北中烟工业有限责任公司	红金龙(硬佳品)	6901028940115	13
34	河北中烟工业公司	钻石(绿石二代)	6901028250764	13

续表

35	河北中烟工业公司	钻石(硬蓝)	6901028080446	12
36	红塔烟草(集团)有限责任公司	玉溪(软)	6901028317177	12
37	红塔烟草(集团)有限责任公司	红塔山(软经典)	6901028315012	13
38	安徽中烟工业公司	黄山(新制皖烟)	6901028225175	13
39	安徽中烟工业公司	黄山(硬一品)	6901028208819	13
40	安徽中烟工业公司	都宝(新)	6901028126014	11
41	浙江中烟工业公司	利群(长嘴)	6901028118828	13
42	浙江中烟工业公司	利群(新版)	6901028118187	13
43	浙江中烟工业公司	大红鹰(软蓝)	6901028120289	13
44	山东中烟工业公司	将军(特纯)	6901028148665	13
45	山东中烟工业公司	泰山(新品)	6901028150194	13
46	广西中烟工业公司	真龙(鸿韵)	6901028014236	13
47	江苏中烟工业公司	苏烟(软金砂)	6901028114455	13
48	江苏中烟工业公司	南京(红)	6901028300063	13
49	江苏中烟工业公司	一品梅(佳品醇)	6901028115520	12
50	川渝中烟工业公司	娇子(时代阳光)	6901028025645	12
51	川渝中烟工业公司	娇子(蓝)	6901028024136	13
52	川渝中烟工业公司	龙凤呈祥(喜庆香烟)	6901028027564	13
53	陕西中烟工业公司	好猫(吉祥)	6901028058827	13
54	陕西中烟工业公司	猴王(金)	6901028058049	13

合型卷烟样品 4 个; 盒标焦油 13 mg 样品 40 个, 12 mg 样品 10 个, 12 mg 以下样品 4 个, 基本涵盖中式卷烟各个类别的代表产品, 具有较好的代表性。

2.2. 评吸方法

评吸小组由经过中式卷烟感官评价方法培训的卷烟调香师班学员及项目组成员组成, 应用中式卷烟风格感官评价方法对卷烟样品进行的感官评价。中式卷烟风格感官评价方法是在香味轮廓评吸法的基础上, 从风格特征(包括 15 个评吸香型、15 个嗅香、4 个口味特征, 共 34 个指标)对卷烟进行评价, 全面涵盖中式卷烟风格感官特征评价的各个部分, 能够对卷烟进行定性分析和定量比较。

2.3. 方法

自 20 世纪 70 年代以来多元数据的图表示法一直是人们所关注的问题, 人们想了不少办法, 这些方法大体上分为两类: 一类是使高维空间的点与平面上的某种图形对应, 这种图形能反映高维数据的某些特点或数据间的某些关系, 我们称为无损作图法; 另一类是在尽可能多地保留原始数据信息的原则下进行降维, 若能使数据维数降至 1 或 2 维, 则可在平面上画图, 我们称这类为有损作图法[4]。

2.3.1. 无损作图法

无损作图法即完全保留原始数据信息, 所有的变量信息都表现在图形中。图中坐标点既可以是原始数据, 也可以是多元统计的分析结果, 常用的无损作图法有统计脸谱法、星形图、平行坐标图等, 无损

作图法仅适合少数样品的可视化比较。

1) 星形图

星形图是雷达图的多元表示形式, 它将每个变量的各个观察单位的数值表示为一个图形, n 个观察单位就有 n 个图, 每个图的每个角表示每个变量。星形图各半径与原指标的对应关系: 从右边起, 水平的半径为第一指针, 逆时针旋转, 星形图的各半径分别对应第二、第三等各个指标, 根据星形图各条半径的长短, 可对样品进行初始分类。

2) 统计脸谱法

统计脸谱图, 也称 Chernoff 脸谱图, 是多元作图的一种经典方法, 能在平面上直观、形象地展示多变量数据的特征, 它有多种变种。其中一种可以绘制最多 36 个变量的数据, 一个变量对应面部的一个特征, 左右半脸各对应 18 个变量。应用此技术有两个关键点, 一是变量的归一化变换(最小值对应 0, 最大值对应 1), 二是面部特征与变量的对照表。利用人类对面部特征细微变化敏感的心理特质, 借助脸谱图, 既可以同时依据多个变量对样品进行相似性或者相异性的综合直观辨别, 也可以识别离群点(例如两个以上变量取值异常生成极端脸形, 又如某个变量取值异常, 对应面部特征有显著变化)。

3) 平行坐标图

平行坐标图是多元作图的又一种经典方法。平行坐标图与脸谱图一样, 突破了笛卡尔坐标系只能表示二维和三维数据的限制, 特别适合展示变量数目超过 3 个的多元数据。其技术原理很简单, 首先对每个变量进行归一化变换, 在横坐标轴上依次排列全部变量, 变量值为纵坐标, 一个样品的各点依次连接为一条折线。离群点看起来是离群的多边形。变量顺序对图形易读性有重要影响, 一种顺序可能比另一种顺序更能反映数据变化的态势。使用颜色表示样品类别可以改善图的易读性。

统计脸谱法、星形图、平行坐标图等无损作图法仅适合少数样品的可视化比较, 当样品数量过多时我们就不能采用无损作图法而选择采用有损作图法。

2.3.2. 有损作图法

当卷烟样品处于高维空间中, 直接采用无损作图法可视化是不现实的, 一个自然地想法是在不严重损失原始数据信息的条件下, 将数据投影到低维空间(2、3 维空间)进行可视化。我们经常用到的有损作图法有主成分分析法、因子分析法和多维标度法等, 本文主要讨论了主成分分析法与多维标度法。

1) 主成分分析[4]

主成分分析是一个常用的数据降维方法, 旨在通过原来变量的少数几个线性组合来解释原来变量绝大多数信息的一种多元统计方法。它把给定的一组相关变量通过线性变换转成另一组不相关的变量, 这些新的变量按照方差依次递减的顺序排列。通常取前几个主成分, 其累计解释的方差比例占 80%-90% 或以上, 以达到降维且不严重损失原数据信息的目的。主成分分析是多维标度法的一种特例, 在多维标度法求古典解时。

$$1) \text{ 由距离阵 } D = (d_{ij}) \text{ 构造 } A = (a_{ij}) = \left(-\frac{1}{2} d_{ij}^2 \right)。$$

$$2) \text{ 令 } B = (b_{ij}), \text{ 使 } b_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}。$$

$$3) \text{ 求 } B \text{ 的特征根 } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n, \text{ 令 } a_{1,k} = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n |\lambda_i|}, \quad a_{2,k} = \frac{\sum_{i=1}^k \lambda_i^2}{\sum_{i=1}^n \lambda_i^2}。$$

这两个量相当于主成分分析中的累计贡献率, 当然我们希望取 k 值不要太大, 而是 $a_{1,k}$ 和 $a_{2,k}$ 比较大。

$$4) \text{ 当 } k \text{ 取定后, 令 } \hat{X} = (\hat{x}_{(1)}, \dots, \hat{x}_{(k)}), \text{ 则 } \hat{X} \text{ 的行向量 } x_{(1)}, \dots, x_{(k)} \text{ 即为欲求的古典解。}$$

由此可见当欧氏距离用作一些特别的代价函数(这些函数能够简化多维标度算法)时, 古典多维标度法

等价于主成分分析。

2) 多维标度法

多维标度法是一种将高维空间的样本投影到低维空间的数据降维方法。多维标度法是以多种研究对象之间某种亲近关系为依据(如距离、相似系数, 亲疏程度的分类情况等), 合理地将研究对象(样品或变量)在低维空间中给出标度或位置, 以便全面而又直观地再现原始各研究对象之间的关系, 同时在此基础上也可按对象点之间距离的远近实现对样品的分类。数学上, 记高维原空间中的样本为 x_1, x_2, \dots, x_n , 维数 p , 样本间两两距离为 d_{ij} , 低维空间中的投影为 y_1, y_2, \dots, y_n , 维数为 $q (q < p)$, 样本间两两距离为 \hat{d}_{ij} , 多维标度法使得投影空间两两距离 \hat{d}_{ij} 尽可能的接近原空间两两距离 d_{ij} 。本文采用多维标度法的 Sammon 非线性映射标准它的步骤[4]如下:

- 1) 已知相似系数矩阵 $D = (d_{ij})$ 。并将其非对角线元素由小到大排列起来:

$$d_{i_1 j_1} \leq d_{i_2 j_2} \leq \dots \leq d_{i_m j_m}, \quad m = \frac{1}{2}n(n-1), \quad i_l < j_l, \quad l = 1, 2, \dots, m$$

- 2) 设 $\hat{X}_{n \times k}$ 是 k 维拟合构造点, 相应的距离阵 $\hat{D} = (\hat{d}_{ij})$, 令 $S^2(\hat{X}) = \frac{\min \sum_{i < j} (d_{ij}^* - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}$ 极小是对一切 $S_k > 10\%$ 进行的, 使上式达到极小的 $\{d_{ij}^*\}$ 称为 \hat{d}_{ij} 对 d_{ij} 的最小二乘单调回归。

如果 $d_{ij}^* \xrightarrow{\text{monotone}} d_{ij}$, 在上式中取 $d_{ij}^* = \hat{d}_{ij} (i < j)$, 这时, $S^2(\hat{X}) = 0$, \hat{X} 是 D 的构造点。

- 3) 若 k 固定, 且能存在一个 \hat{X}_0 , 使得 $S^2(\hat{X}_0) = \min_{\hat{X}_{n \times k}} S^2(\hat{X}) = S_k$ 则称为 S_k 为 k 维最佳拟合构造点。

4) 由于 S_k (也称压力指数)是 k 的单调下降序列, 取 k 使 S_k 适当小。例如 $S_k < 5\%$ 最好, $S_k > 10\%$ 较差。

本研究中以中式卷烟风格感官评价方法[5]-[8]为例演示多维标度法, 但该方法也适用于其他卷烟感官评价方法。针对评吸数据, 分别采用主成分分析和多维标度法进行数据可视化。计算过程采用 MATLAB 2011 完成。

3. 结果与讨论

3.1. 星形图结果分析

图 1 是根据卷烟风格特征数据所做的星形图。共有 54 个样本, 每个星形图的角度表示一个变量。因为各个样品图形差异并不明显, 从图 1 中很难看出那几个样品相似度高, 只有在样本较少的情况下我们才能方便的对各个样本进行比较。由此可知星行图并不适合对高维数据可视化。

3.2. 统计脸谱法

图 2 是根据卷烟风格特征数据所做的统计脸谱图。脸谱图生动、直观, 能够非常形象的表达样本之间的差异。每一个脸谱图代表一个样品, 从图 2 我们可以看出都宝, 中南海, 金桥和林海灵芝 4 个样品与其余 50 样品图形差异较大, 受样品数量影响很难在较多样品的情况下看出其余几个样品之间的不相似度。

3.3. 平行坐标图

图 3 是根据卷烟风格特征数据所做的平行坐标图。每一条折线代表一个样本, 从图 3 中我们很容易发现有 4 个样品与其他 50 个样品有明显的差异性, 但是由于样品数量太多我们凭肉眼观察要从图例中找到所对应 4 个差异样品比较困难, 因为是 54 个样品, 每种样品对应一种颜色, 因样品太多我们很难分别



Figure 1. 54 sample star diagrams
图 1. 54 个样品星形图

出哪种样品具体对应的是哪条折线和哪种颜色, 所以我们只能客观的知道至少有 4 种样品与其余 50 种样品差异较大。

3.4. 主成分分析法结果分析

对样本进行主成分分析, 图 4 采用主成分分析将 34 维原始数据投影到二维空间的结果。由下面图 4 可知:

- 1) 都宝, 中南海, 金桥和林海灵芝 4 个样品与其余 50 样品不相似度较大。
- 2) 五叶神和金圣(硬)与其余 52 个样品不相似度较大。

投影维数的确定

我们得到前几个主成分解释的方差比例比较小(前 2、3 个主成分分别解释 44.2%, 57.5%), 见图 5。也就是说, 降维后原始数据的信息会有严重损失。事实上, 由图 5 可知, 在主成分个数为 10 的地方有明显的转折点, 建议取 10 个主成分, 也就是取 10 维。这意味着利用主成分分析来对评吸数据可视化(投影到 2、3 维)不太合适。不过, 主成分分析结果也揭示本项目所采用的评吸指标之间相关度不高, 也就是



Figure 2. 54 sample Facebook
图 2. 54 个样品脸谱图

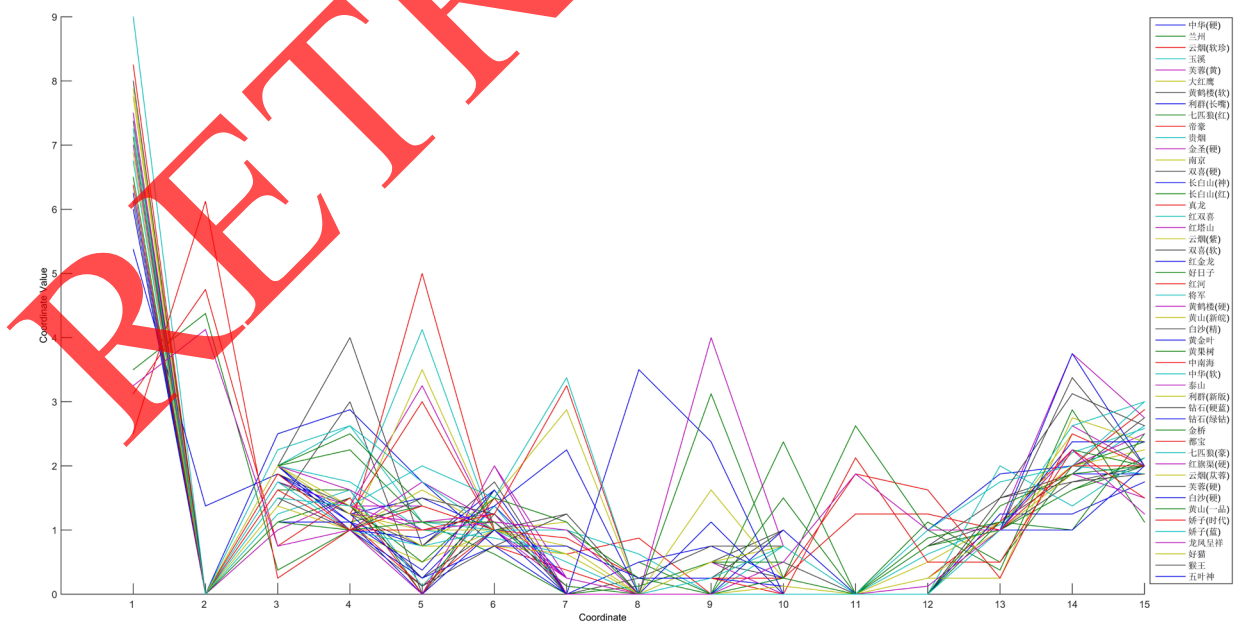


Figure 3. 54 sample parallel coordinates
图 3. 54 个样品平行坐标图

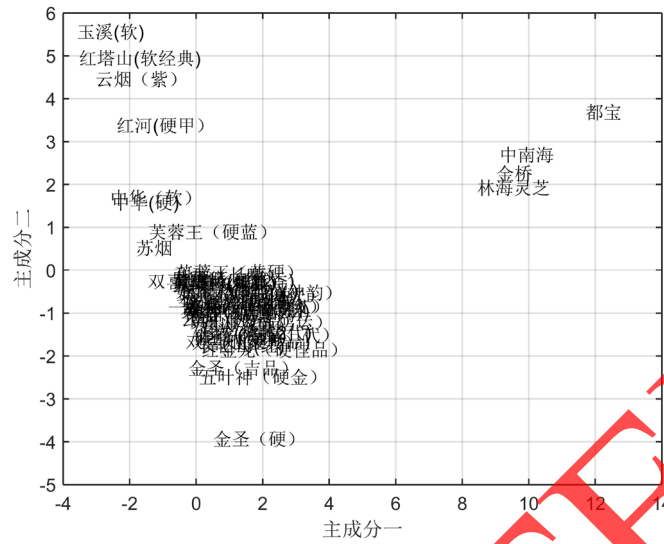


Figure 4. Principal component analysis of 2 dimensional scatter plot
图 4. 主成分分析 2 维的散点图

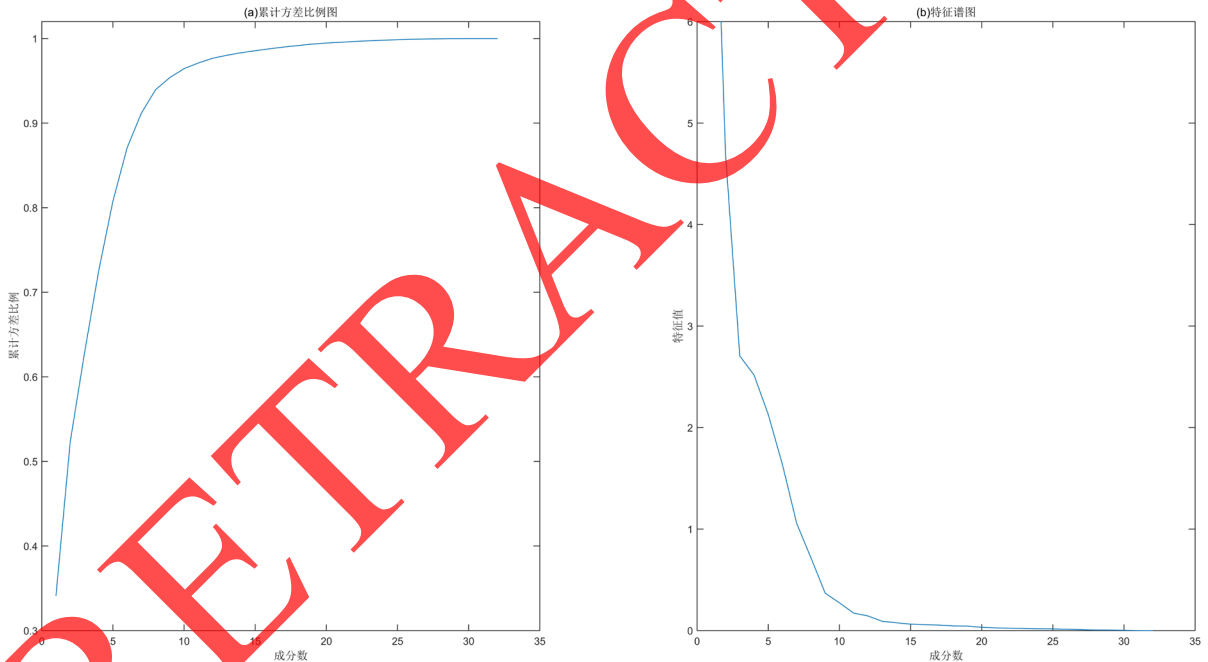


Figure 5. Principal component analysis of the cumulative variance and eigenvalue spectra picture
图 5. 主成分分析的累积方差和特征值谱图

说指标间重叠度较低, 较为科学。

3.5. 多维标度法

3.5.1. 多维标度法结果分析

鉴于对主成分分析方法的分析, 我们对样本采用多维标度法, 因该法能够最大程度保留我们所想要的原始空间中两两样品的不相似度。多维标度法有几种不同的映射标准, 不同的标准可能会得到不同投影结果。本文研究 Sammon 非线性映射标准, 侧重于保留原始空间中小的两两不相似度。图 6 显示了采

用 Sammon 非线性映射标准将 34 维原始数据投影到二维空间的结果。由图 6 可知:

- 1) 中南海, 金桥, 都宝, 林海灵芝 4 个样品与其余 50 样品不相似度较大。
- 2) 五叶神与其余 53 个样品不相似度较大。

图 7 可以帮助我们了解卷烟样品的两两不相似度。

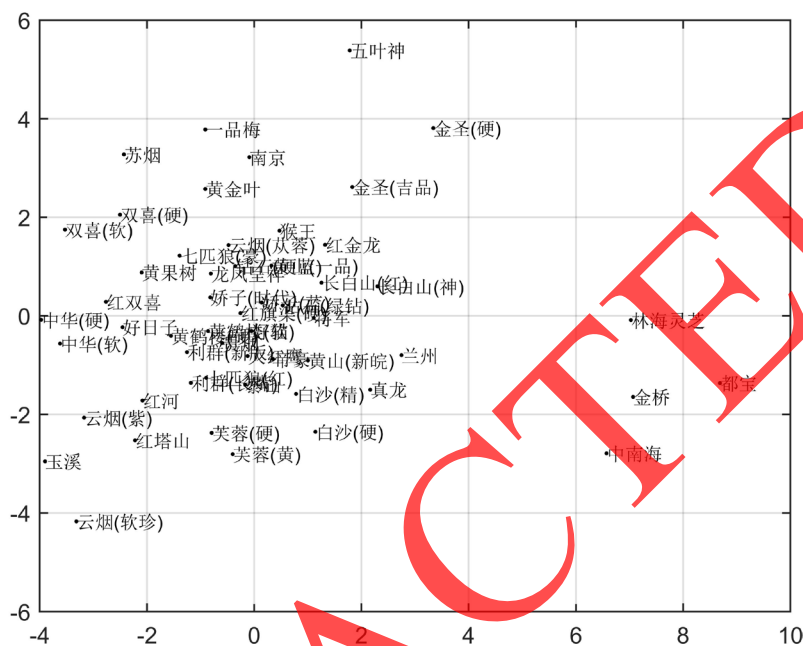


Figure 6. The multidimensional scaling method is used for the projection of the Sammon nonlinear mapping standard to the 2 dimensional scatter plot
图 6. 多维标度法采用 Sammon 非线性映射标准投影到 2 维的散点图

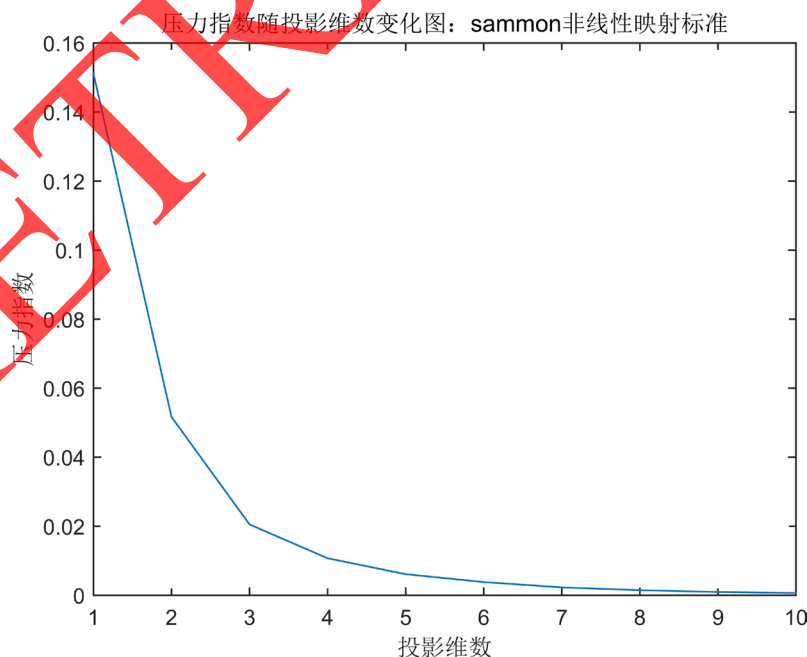


Figure 7. The pressure exponent with the projection dimension change map
图 7. 压力指数随投影维数变化图

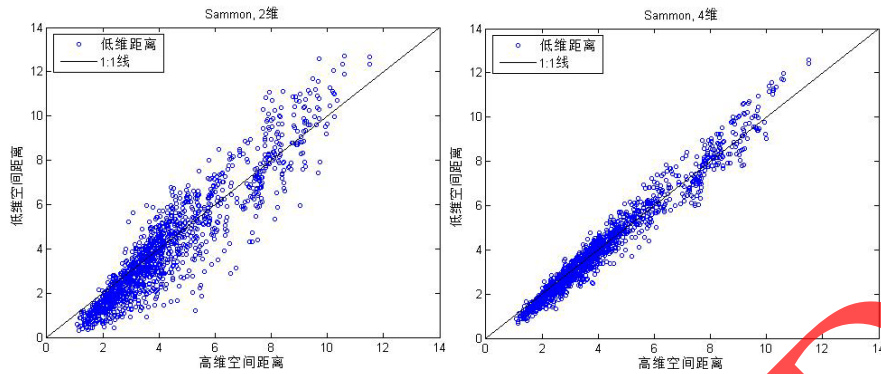


Figure 8. Shepard chart
图 8. Shepard 图

3.5.2. 投影维数的确定

类似于主成分分析, 多维标度法作为一个数据降维方法需要决定数据投影到几维比较合适。一个常用的方法是作出多维标度法的目标函数值(压力指数)随投影维数的变化图, 有明显转折点的地方即可选为投影维数, 因为该转折点之后, 随着维数的增加, 对目标函数的贡献变小, 也就是重要程度变低, 所以可以考虑舍弃。图 7 为多维标度法采用 Sammon 映射标准的压力指数随投影维数(1-10)的变化图。由图 7 可以观察到, Sammon 标准取投影维数 4, 该图的转折点较明显。

图 8 为采用 Sammon 映射标准分别投影到 2 维和 4 维的 Shepard 图, Shepard 图是一个散点图, 纵轴表示低维空间中的距离, 横轴表示高维空间中的距离, 如果低维空间较好的保留原空间的不相似度, 那么所有的点就在 1:1 线附近。我们发现在 4 维时该图能够反映降维后保留原始两两不相似度的好坏程度。

由图 8 可知: 1、Sammon 标准保留原空间小的不相似度较好。2、Sammon 4 维 Shepard 图比 2 维 Shepard 图更靠近 1:1 线, 由此验证 Sammon 标准在 4 维空间能更好的保留原始空间中两两样品的不相似度。

4. 结论

统计脸谱法、星形图、平行坐标图等无损作图法在样品比较多时很难对各个样品进行比较, 因此无损作图法仅适合少数样品的可视化比较。主成分分析需要取 10 个主成分才能保证降维后原始数据的信息不会严重损失, 这意味着利用主成分分析来对评吸数据可视化(投影到 2、3 维)也不太合适。多维标度法使用 Sammon 标准观察投影图只需要在 4 维就可以对评吸数据可视化, 可见多维标度法适合对评吸数据可视化。因此, 当主成分分析不可靠时, 多维标度法是一个更佳选择。

参考文献 (References)

- [1] 何晓群. 多元统计分析[M]. 第 2 版, 北京: 中国人民大学出版社, 2008.
- [2] 杨锦忠, 宋希云, 等. 多元统计分析及其在烟草学中的应用[J]. 中国烟草学报, 2014(10): 134-138.
- [3] Izenman, A.J. (2008) Modern Multivariate Statistical Techniques. *Multidimensional Scaling and Distance Geometry*.
- [4] 王斌会. 多元统计分析及 R 语言建模[M]. 广州: 暨南大学出版社, 2009: 227.
- [5] 武怡, 张虹, 赵震毅, 等. 中式卷烟风格特征感官评价方法的建立与应用[J]. 中国烟草, 2008(2): 69-71.
- [6] 邵干辉, 袁庆钊, 温东奇, 等. 国内国际品牌卷烟香味特征和致香成分比较分析[J]. 广东科技, 2008(20): 43-46.
- [7] 翁昔阳, 刘艺, 许建铭, 等. 香味轮廓分析法在卷烟评香中的应用[C]//中国烟草学会工业专业委员会烟草化学学术研讨会论文集. 海口: 中国烟草学会, 2005: 468-472.
- [8] 武怡, 曾晓鹰, 朱保昆, 等. 中式卷烟风格感官评价方法区域适应性分析[J]. 烟草科技, 2012(9): 5-9.