网络首发时间：2019-07-23 13:15:11
网络首发地址：http://kns.cnki.net/kcms/detail/42.1009.C.20190722.1527.017.html

DOI:10.13546/j.cnki.tjyjc.2019.14.019

低维逐次投影寻踪模型及其应用

于晓虹[a]，楼文高[b]
（上海商学院a.东方财富传媒与管理学院；b.信息与计算机学院，上海 200235）

摘　要：文章建立了一维逐次投影寻踪（SPPC）模型的目标函数和约束条件，提出采用矢量合成法构建低维逐次PPC（LDSPPC）模型，确定各SPPC模型的权重分配原则，推导出LDSPPC模型最佳投影向量、样本投影值等计算公式。针对六个供应商9个评价指标的样本数据，建立了LDSPPC模型，求得各个评价指标的客观权重和六个供应商的得分。结果表明：市场占有率对供应商优劣影响最大，其次是产品合格率和净资产收益率；六个供应商中，供应商S4最优，S1最劣。LDSPPC模型与人类进行综合评价的思维方式基本一致，能从样本数据中挖掘出更多的信息，具有较好的客观性和合理性，是一种供应商选择与评估的新方法。

关键词：逐次；一维投影寻踪模型；供应商；选择与评估
中图分类号：TP391　　**文献标识码**：A　　**文章编号**：1002-6487（2019）14-0083-04

0　引言

1974年Friedman等[1]提出了将高维样本数据投影到低维子空间上的投影寻踪聚类（简称PPC）原理，通过研究不同投影方向（空间角度）上样本投影点的分布规律、结构特性，以确定最优（也称为“感兴趣”）投影方向。由于求最优解非常困难，目前最常用的是一维PPC模型，其目标函数是“使样本投影值$z(i)$的标准差S_z与局部密度值D_z的乘积最大化”，建模基本思想是“使样本点在整体上尽可能分散，并形成若干类（团），类与类之间尽可能分开，而各类内的样本点则应尽可能密集”。PPC建模基本思想与人类进行综合评价的思维方式基本一致，而且可同时求得评价指标的客观权重和研究对象（样本）的综合得分，是一种较理想的综合评价方法，在学术界获得了广泛的应用[2-6]。根据上述PPC建模基本思想，不同学者提出了多种不同目标函数型式，但仍以Friedman等[1]提出的PPC模型效果最好。

为了从样本数据中挖掘出更多的有效信息，Friedman等[1]提出可逐次建立第2、第3和第4维的SPPC模型。由于求解SPPC模型目标函数的最优解非常困难，本文提出建立低维（1—4维）逐次PPC（简称LDSPPC）模型的基本原理（已申请国家发明专利）；建立逐次PPC模型的目标函数与约束条件；提出判定最优化过程是否求得真正全局最优解的判定准则和方法，从而确保求得各SPPC模型真正的全局最优解；提出采用矢量合成法将多个SPPC模型构建成综合的LDSPPC模型；将LDSPPC模型应用于供应商选择与评估研究，进行实证建模。结果表明，LDSPPC模型能比一维PPC模型挖掘出更多的样本数据信息，是一种理想的综合评价新方法。

1　低维逐次PPC模型的原理

设样本原始数据为$x^*(i,j)$ $(i=1,2,\cdots,n;j=1,2,\cdots,p)$（$n$、$p$分别为样本个数和变量维数）。

（1）对样本原始数据进行无量纲化预处理，令无量纲化数据为$x(i,j)$。

（2）计算各个样本的投影值。

假设已求得第m维SPPC模型的最佳投影向量$\vec{a}_m=[a_m(1),a_m(2),\cdots,a_m(p)]$，则其样本投影值（评价值）为$z_m(i)=\sum_{j=1}^{p}a_m(j)*x(i,j)$。假设共建立了M维SPPC模型，$M$小于等于4。

（3）构建低维（1—M维）SPPC模型的目标函数与约束条件。

本文以聚类效果最好的、由Friedman等[1]提出的一维PPC模型为基[1-6]，建立LDSPPC模型。则提出第m维SPPC模型的目标函数为：

$$\begin{aligned}&Q_m(\vec{a}_m)=\max(S_{z,m}*D_{z,m})\\&s.t.\sum_{j=1}^{p}a_m^2(j)=1,1\geq a_m(j)\geq-1\\&\sum_{j=1}^{p}a_t(j)*a_{t+1}(j)=0,(t=1,2,\ldots,m-1)\end{aligned}\tag{1}$$

其中，第m维SPPC模型样本投影值标准差$S_{z,m}=\sqrt{\left\{\sum_{i=1}^{n}\left[z_m(i)-\overline{z_m}\right]^2\right\}/(n-1)}$，其值越大表示样本投影点整体

基金项目：全国统计科学研究一般项目（2016LY93）；上海高校知识服务平台“上海商贸服务业知识服务中心”项目（ZF1226）
作者简介：于晓虹（1978—），女，安徽合肥人，硕士，讲师，研究方向：金融工程、供应链管理。
（通讯作者）楼文高（1964—），男，浙江杭州人，博士，教授，研究方向：人工神经网络、投影寻踪理论。

上越分散，局部密度值 $D_{z,m}=\sum_{i=1}^{n}\sum_{k=1}^{n}[R_m-r_m(i,k)]\cdot u[R_m-r_m(i,k)]$，其值越大表示各个类内的样本投影点越密集，$\overline{z_m}$ 为 $z_m(i)$ 的均值，$r_m(i,k)=|z_m(i)-z_m(k)|$ 表示样本 i 与 k 投影点之间的距离，$r_{\max,m}$ 为 $r_m(i,k)$ 的最大值，$u(t)$ 为单位阶跃函数，当 $t\geq 0$ 时为1，否则为0，R_m 为窗宽半径，其合理取值范围为 $r_{\max,m}/5\leq R_m\leq r_{\max,m}/3$，本文取 $R_m=r_{\max,m}/5$ 进行建模（取 $R_m=r_{\max,m}/3$ 时结果基本相同）。对于一维PPC模型，根据样本投影值 $z_1(i)$ 的大小就可以进行优劣排序和分类研究了。

（4）将上述M维SPPC模型矢量合成为LDSPPC模型。

Friedman等[1]虽然提出了可逐次建立一维PPC模型的设想，但没有提出如何将SPPC模型合成为一个综合模型，导致无法利用上述SPPC模型对样本进行排序研究。本文认为，上述第1、第2、…、第M维最佳投影向量 $\vec{a}_1$、$\vec{a}_2$、…、$\vec{a}_M$ 是相互垂直的，如果对多个SPPC模型的样本投影值进行简单加权求和处理，再进行排序研究是缺乏理论依据和数学意义的。而且，PPC建模基本思想是将样本数据投影到某个“感兴趣”的方向上，其合成模型也应该具有这个“投影”特性。建立SPPC模型的目标函数是使 $Q_m(\vec{a}_m)$ 最大化，即 $Q_m(\vec{a}_m)$ 表示了第 m 维SPPC模型的相对重要性，即可根据 $Q_m(\vec{a}_m)$ 的大小为各SPPC模型分配权重。采用矢量合成法将多个SPPC模型构建成LDSPPC模型，仍然是在某个方向上的投影。

令第 m 维SPPC模型的目标函数值为 $Q_m(\vec{a}_m)$，则根据矢量合成法原理其分配权重 ω_m 为：$\omega_m=\dfrac{Q_m(\vec{a}_m)}{\sqrt{\sum_{h=1}^{M}Q_h^2(\vec{a}_h)}}$ $(m=1,2,\cdots,M)$，则LDSPPC模型的最佳投影向量 $\vec{a}_z=\{a_z(1),a_z(2),\cdots,a_z(p)\}$，满足 $\sum_{j=1}^{p}a_z^2(j)=\sum_{j=1}^{p}\left[\sum_{h=1}^{M}\omega_h a_h(j)\right]^2=1$。从而有LDSPPC模型的样本综合投影值 $z_z(i)=\sum_{j=1}^{p}a_z(j)*x(i,j)$。

（5）确定需要建立SPPC模型的维数。

一般不超过4个，一般要求第 M 维SPPC模型的权重占比小于等于0.20，因为 $Q_1\geq Q_2\geq\cdots\geq Q_M$，最不利的情况是 $\dfrac{Q_M(\vec{a}_M)}{Q_1(\vec{a}_1)+(M-1)Q_M(\vec{a}_M)}\leq 0.20$，则 $M=2,3,4$ 是分别要求 $Q_2(\vec{a}_2)\leq 0.25Q_1(\vec{a}_1)$、$Q_3(\vec{a}_3)\leq 0.33Q_1(\vec{a}_1)$ 和 $Q_4(\vec{a}_4)\leq 0.50Q_1(\vec{a}_1)$，实际占比必定小于等于0.20。

（6）根据LDSPPC模型最佳投影向量系数 $a_z(j)$ 的大小，可以确定各个评价指标的重要性及其排序与分类结果，根据研究对象（样本）投影值 $z_z(i)$ 的大小，可以得到研究对象的优劣排序与分类结果。

（7）求解SPPC模型的真正全局最优解。

SPPC模型为同时含有等式和不等式约束的高维非线性最优化问题，求解非常困难。本文通过对几十种群智能最优化算法的比较研究发现，群搜索（Group search optimization，简称GSO）[7]算法具有较好的全局搜索能力和较快收敛速度。为此，用基于GSO的Matlab最优化程序求解各SPPC模型的最佳投影向量 $\vec{a}_m$ 等。根据楼文高等（2015），通过改变一半指标的归一化方式，如果其权重变为相反数而目标函数值 $Q_m(\vec{a}_m)$、$S_{z,m}$、$D_{z,m}$ 等保持不变，则可以判断最优化过程已经求得了各SPPC模型的真正全局最优解。

2 建立供应商选择与评估的LDSPPC模型

2.1 供应商选择与评估方法

供应商选择与评估是一个典型的多属性（因素）综合评价问题，主要评价方法（模型）有各种（改进型）TOPSIS、灰色关联法（GCM）等。马丽娟[8]提出的供应商选择与评估案例被多位学者引用，如孙晓东等[9]、王硕等[10]、王先甲等[11]和张莉等[12]。王先甲等[11]认为当评价指标之间相关性较高时，基于欧氏距离TOPSIS的结果不一定可靠，张莉等[12]认为文献[11]采用的马氏距离要求样本数量（研究对象、备选方案数）必须多于评价指标个数，绝大多数供应商选择与评估问题不满足这个条件，提出用基于余弦相似性的TOPSIS法进行评价，得到与文献[10]基本一致的结果。经本文多次建模发现，文献[10]没有求得真正的全局最优解，其结果是错误的（详见讨论部分）。因此，供应商选择与评估问题仍然是需要进一步深入研究的课题。

本文将LDSPPC模型应用于供应商选择与评估，可同时求得各个评价指标权重和研究对象的评价结果，以期得到更合理、客观和有效的结果，以及提供一种新方法。

2.2 供应商选择与评估实例

文献[9,10,12]采用的六个供应商的评价指标数据如表1所示。

表1　　供应商S1—S6的评价指标数据

供应商	x_1^*	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
S1	335	3.2	15	0.8	0.12	230	0.10	0.82	0.13
S2	268	1.4	37	0.92	0.25	130	0.08	0.96	0.15
S3	304	1.9	22	0.99	0.09	220	0.14	0.99	0.20
S4	270	2.0	16	0.98	0.35	180	0.12	0.96	0.21
S5	310	0.8	26	0.86	0.20	150	0.15	0.8	0.12
S6	303	2.7	10	0.95	0.19	170	0.16	0.91	0.19

注：评价指标 x_1 至 x_9 分别为产品价格（元）、售后服务（小时）、地理位置（公里）、产品合格率、新产品开发率、供应能力（件）、净资产收益率、准时交货率和市场占有率，其中 x_1 至 x_3 为越小越好的成本型指标，其他为越大越好的效益型指标。

将表1数据进行极差归一化预处理，导入本文编制的LDSPPC程序，得到了第1维SPPC模型的真正全局最优解：第1维最佳投影向量 $\vec{a}_1$、$Q_1(\vec{a}_1)$、$S_{z,1}$、$D_{z,1}$、R_1 和 $r_{1,\max}$ 如下页表2所示，供应商S1至S6的第1维样本投影值 $\{z_1(1),z_1(2),\ldots,z_1(6)\}$ 也列于表2中。从供应商S1至S6的第1维投影值可以看出，供应商S4最优，供应商S1最差，S2等4个供应商的投影值相等，无法对其进行排序，这

个结果真正体现了PPC模型“使局部投影点尽可能密集”的特性——有几个样本的投影值是相等的。

为了进一步挖掘样本数据信息，应继续建立第2维SPPC模型。再次调用LDSPPC程序求得第2维模型的真正全局最优解：第2维最佳投影向量系数 $\{a_2(1), a_2(2), \cdots, a_2(9)\}$ 、$Q_2(\vec{a}_2)$ 、$S_{z,2}$ 、$D_{z,2}$ 、R_2 、$r_{2,\max}$ 、供应商S1至S6的投影值 $\{z_2(1), z_2(2), \cdots, z_2(6)\}$ 如表2所示。从第2维SPPC模型投影值可以看出，供应商S4最优，供应商S5最差，供应商S2、S3和S6的投影值相等。根据第1维、第2维投影值，仍然无法对供应商S2、S3和S6进行排序。

由于 $Q_2(\vec{a}_2)>0.25Q_1(\vec{a}_1)$ ，$Q_3(\vec{a}_3)>0.33Q_1(\vec{a}_1)$ ，$Q_4(\vec{a}_4)<0.5Q_1(\vec{a}_1)$ ，所以本文共建立第4维SPPC模型。第3、第4维SPPC模型的最佳投影向量系数、目标函数值等也列于表2中。

表2　各维SPPC模型及其LDSPPC模型最佳投影向量系数、样本投影值等参数

参数	第1维	第2维	第3维	第4维	LDSPPC
$a_m(1)$	0.2302	0.3621	–0.2423	–0.1833	0.1940
$a_m(2)$	0.4649	–0.3490	–0.2219	0.1973	0.1255
$a_m(3)$	–0.1152	0.2181	0.6268	–0.1332	0.2461
$a_m(4)$	0.4202	0.2640	0.0121	0.2614	0.5047
$a_m(5)$	0.0646	0.3975	0.1219	–0.4222	0.1788
$a_m(6)$	–0.4161	0.0436	0.2015	0.6869	–0.0089
$a_m(7)$	0.5546	–0.2905	0.5523	0.0708	0.5016
$a_m(8)$	0.1405	0.4011	–0.3100	0.3325	0.2639
$a_m(9)$	0.1906	0.4740	0.2061	0.2751	0.5300
$z_m(1)$	–0.3288[6]	0.2897	0.8568	0.6126	0.4237[6]
$z_m(2)$	1.0651	1.0077	–0.5160[4]	0.2413	1.1152[4]
$z_m(3)$	1.0651	1.0077	0.5995	1.4575[3]	1.9016[2]
$z_m(4)$	1.2063[1]	1.6826	0.5995	0.5763	2.0972[1]
$z_m(5)$	1.0651	–0.1187[5]	0.5246	0.1775	0.9697[5]
$z_m(6)$	1.0651	1.0077	1.1371[2]	0.6160	1.8945[3]
$S_{z,m}$	0.5833	0.6343	0.5620	0.4566	0.6636
$D_{z,m}$	6.8529	4.3231	4.0615	3.2975	3.5840
$Q_m(\vec{a}_m)$	3.9976	2.7422	2.2825	1.5055	2.3783
R_m	0.3070	0.3603	0.3306	0.2560	0.3347
$r_{m,\max}$	1.5351	1.8013	1.6531	1.2800	1.6735

注：1、2、…、6表示在该SPPC模型中该供应商的排序。

2.3　各维SPPC模型的矢量合成

根据前述原理，各维SPPC模型的权重 $\omega_m=Q_m(\vec{a}_m)/\sqrt{\sum_{h=1}^{4}Q_h^2(\vec{a}_h)}$ ，可以求得第1、第2、第3和第4维SPPC模型的分配权重分别为0.7183、0.4927、0.4101和0.2705，显然，第1维SPPC模型在矢量合成后的LDSPPC模型中占显著的主导地位，第2维与第3维相差较小，明显大于第4维的作用。经矢量合成得到LDSPPC模型的最佳投影向量 $\vec{a}_z=\{a_z(1), a_z(2), \cdots, a_z(9)\}=(0.1940,0.1255,0.2461,0.5047,0.1788,-0.0089,0.5016,0.2639,0.5300)$，如表2所示。可以看出，$x_9$ 的权重最大，其次是 x_4 、x_7 ，而且这3个指标的权重显著大于其他指标的权重；然后按照权重大小依次是 $x_8>x_3>x_1>x_5>x_2>x_6$ ，其中 x_6 的权重几乎等于0（略小于0），可以删除。

根据矢量合成法原理进而求得六个供应商LDSPPC模型的样本投影值 $\{z_z(1), z_z(2), \cdots, z_z(6)\}=(0.4237, 1.1152, 1.9016, 2.0972, 0.9697, 1.8945)$。显然，供应商S4最优，其次是S3，供应商S1最劣，供应商优劣排序为S4>S3>S6>S2>S5>S1，而且S3和S6相差不大。

3　结果与讨论

3.1　矢量合成LDSPPC模型有关问题的讨论

（1）采用目标函数值大小分配各维SPPC模型权重和采用矢量合成法构建LDSPPC模型的合理性分析。因为建立各SPPC模型时就是使其目标函数值最大化，其大小表明了在LDSPPC模型中的作用大小，因此，用其分配权重是合理和可行的。

（2）必须用楼文高等（2015）提出的定理1、2和3来判断最优化过程是否求得了各为SPPC模型的真正全局最优解，否则建模结果必定是错误的。一旦第 m 维SPPC模型的结果是错误的，其后续所有模型的结果都必定是错误的。所有群智能最优化算法，都不能保证每次运行都一定能求得真正的全局最优解。

3.2　建立LDSPPC模型的必要性分析

从第1—4维SPPC模型的权重可知，第1维SPPC模型仅挖掘出38%的样本信息，第2、第3和第4维分别占比26%、22%和14%。因此，对于本文研究的供应商选择与评估问题，必须采用LDSPPC模型，否则，仅建立一维PPC模型并不能充分反映六个供应商的真实情况，也不能区分其中4个供应商的优劣。

3.3　六个供应商优劣排序及其各个评价指标的重要性分析和排序

对供应商选择与评估来说，不仅要求得供应商的优劣排序，更应该求得评价指标的重要性排序，对后续选择其他供应商更具有指导意义。

从LDSPPC模型可知，9个评价指标的归一化权重分别为0.0760、0.0491、0.0964、0.1977、0.0700、0.0035、0.1964、0.1033和0.2076，其中市场占有率 (x_9) 对供应商最重要，归一化权重占比为20.8%，其次是产品合格率 (x_4) ，权重占比为19.8%，再次是净资产收益率 (x_7) ，权重占比为19.6%，其他指标的权重占比多数小于10%，明显小于上述三个指标。其中 x_9 、x_4 和 x_7 为三个最重要指标，重要指标有 x_8 、x_3 、x_1 和 x_5 ，供应能力指标 (x_6) 归一化权重很小，可以删除。

从LDSPPC模型的六个供应商得分可知，供应商优劣排序为S4>S3>S6>S2>S5>S1，供应商S4最优，S1最劣。

3.4　不同评价方法结果的比较

文献[9,10,12]采用不同评价方法研究了本例，供应商优劣排序结果如表3所示。从表3可以看出：（1）供应商S4都是最优的，S1都是最劣的，其他供应商排名不一致；（2）文献[10]没有求得真正的全局最优解，根据其给出的评价指标权重、偏好系数，求得其目标函数值只有0.056167。经多次最优化试验，求得真正的全局最优解：

评价指标权重为0.209502、0、0、0、0.012134、0、0、0.778364、0，偏好系数a_1 = 1.00000、a_2 = 0、b_1 =1.00000、b_2 = 0，因为a_1和b_1等于1，分辨系数x可以取任何值，目标函数值为0.184026，六个供应商S1至S6的投影值分别为0.1016、0.8477、0.8505、0.8505、0.1016、0.5716，这个结果体现了PPC模型“局部尽可能密集”的特点——总有几个样本的投影值是相等的，其优劣排序如表3所示，无法实现完全排序。

由于供应商数量少于评价指标个数，不能用文献[12]的方法建模，由于评价指标之间相关性较高，不能用文献[8–10]的方法建模，其结果并没有可比性。本文结果与文献[12]结果存在一定差异，根据逐次建立的SPPC模型，本文方法可以很清晰地看出判定供应商优劣的次序和过程，既形象又明确具体，同时得到了各个评价指标的权重及其重要性排序，对后续选择其他供应商具有很好的指导意义，是一种理想的综合评价方法。而且文献[12]无法判定各个评价指标的重要性，对今后如何选择供应商没有指导意义。

表3　　不同评价方法六个供应商的优劣排序结果

评价方法	S1	S2	S3	S4	S5	S6
一维PPC	6	2	2	1	2	2
文献[9]*	6	5	4	1	2	3
文献[10]	6	5	3	1	4	2
文献[10]+	6	3	1	1	6	4
文献[12]	6	5	4	1	3	2
LDSPPC	6	4	2	1	5	3

注：*决策者偏好不同，S2、S3的排序有可能改变，最优供应商排序第1，最劣第6。+是采用文献[11]的评价方法和数据求得真正全局最优解时的结果。

4　结束语

本文建立了各维SPPC模型的目标函数与约束条件，提出根据各SPPC模型的目标函数值大小分配其权重和停止建立SPPC模型的原则，提出采用矢量合成法构建LDSPPC模型，以保障其仍然具有PPC模型的空间“投影”特性，推导得到LDSPPC模型最佳投影向量、样本投影值等计算公式。与一维PPC模型相比，LDSPPC模型能从样本数据中挖掘出更充分的信息，提高建模结果的真实性和可靠性，具有更好的客观性和有效性。

采用LDSPPC模型进行供应商选择与评估，不仅数学意义和物理意义更清晰，而且可以同时求得各个评价指标的权重和各个供应商的优劣，并对各个评价指标进行重要性排序与分类，实现供应商优劣排序与分类，属于客观建模方法，评价结果更客观、合理和有效。

参考文献：

[1]Friedman J H, Tukey J W. A Projection Pursuit Algorithm for Exploratory Data Analysis [J]. IEEE Transactions on Computers, 1974, 23(9).

[2]付强, 赵小勇. 投影寻踪模型原理及其应用[M]. 北京:科学出版社, 2006.

[3]于晓虹, 楼文高, 余秀荣. 中国省际普惠金融发展水平综合评价与实证研究[J]. 金融论坛, 2016,(5).

[4]楼文高, 吴晓伟. 区域流通业竞争力投影寻踪建模及实证研究 [J]. 中国流通经济, 2010,(10).

[5]虞玉华，楼文高.体育类期刊学术水平综合评价与实证研究——基于决策者偏好的投影寻踪建模技术[J]. 北京体育大学学报, 2015, 38(12).

[6]楼文高, 干瑞娟, 李坦. 基于投影寻踪模型的图书馆成效(绩效)评估研究[J]. 图书情报工作，2017, 61(9).

[7]张雯雰, 刘华艳. 改进的群搜索优化算法在MATLAB中的实现[J]. 电脑与信息技术, 2010, 18(3).

[8]马丽娟.基于供应链管理的供应商选择问题初探[J]. 工业工程与管理，2002,(6).

[9]孙晓东, 焦玥, 胡劲松.基于灰色关联度和理想解法的决策方法研究[J]. 中国管理科学，2005,13(4).

[10]王 硕, 杨善林, 胡笑旋.基于投影寻踪的组合评价方法研究[J]. 中国工程科学，2008，10(8).

[11]王先甲, 汪磊. 基于马氏距离的改进型TOPSIS在供应商选择中的应用[J]. 控制与决策, 2012, 27(10).

[12]张莉，夏佩佩，李凡长. 基于余弦相似性的供应商选择方法[J]. 山东大学学报(工学版), 2017, 47(1).

（责任编辑/浩　天）