因子分析算法开发文档

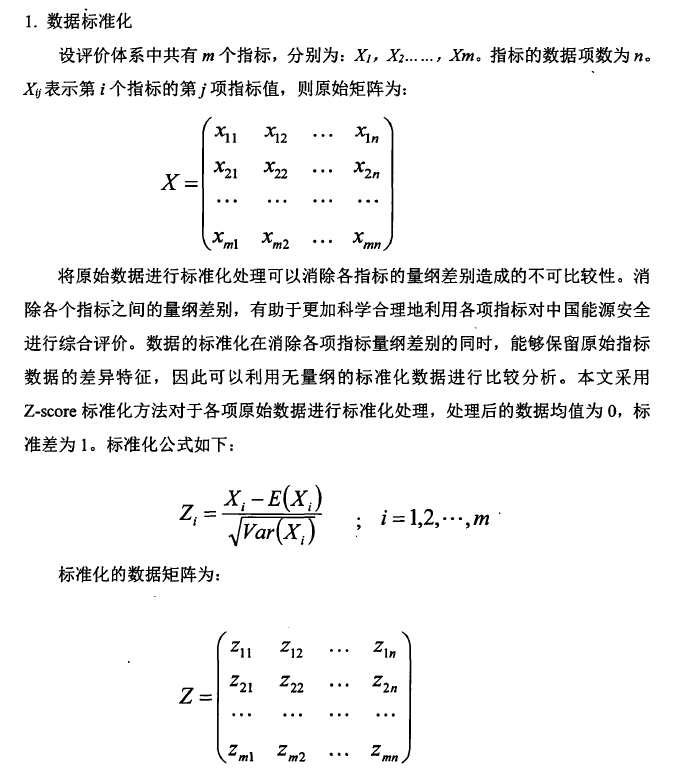
# 1 算法原理

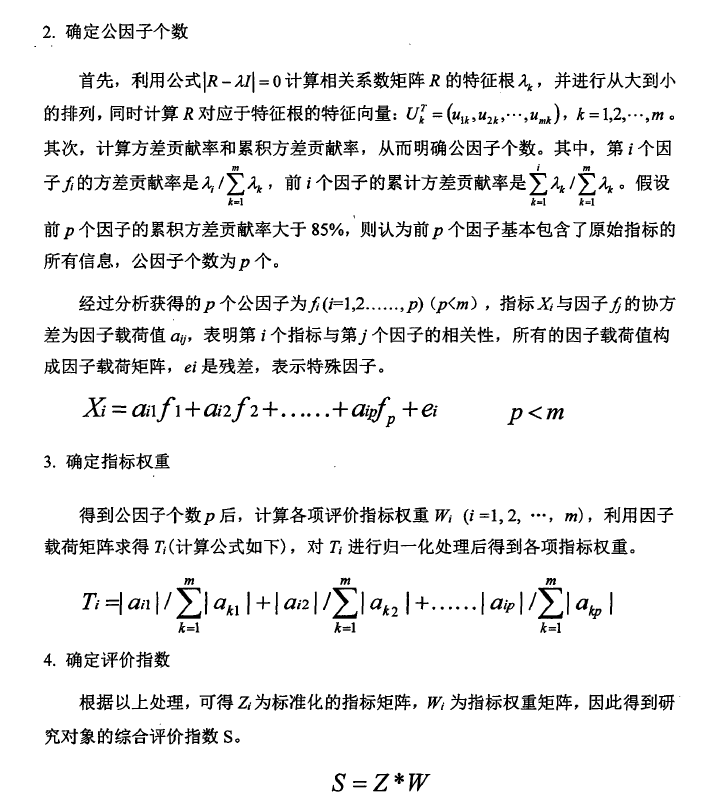
## 1.1 算法概述

因子分析是指研究从变量群中提取共性因子的统计技术。最早由英国心理学家C.E.斯皮尔曼提出。他发现学生的各科成绩之间存在着一定的相关性，一科成绩好的学生，往往其他各科成绩也比较好，从而推想是否存在某些潜在的共性因子，或称某些一般智力条件影响着学生的学习成绩。因子分析可在许多变量中找出隐藏的具有代表性的因子。将相同本质的变量归入一个因子，可减少变量的数目，还可检验变量间关系的假设。

## 1.2 算法组成与步骤

因子分析的具体步骤:





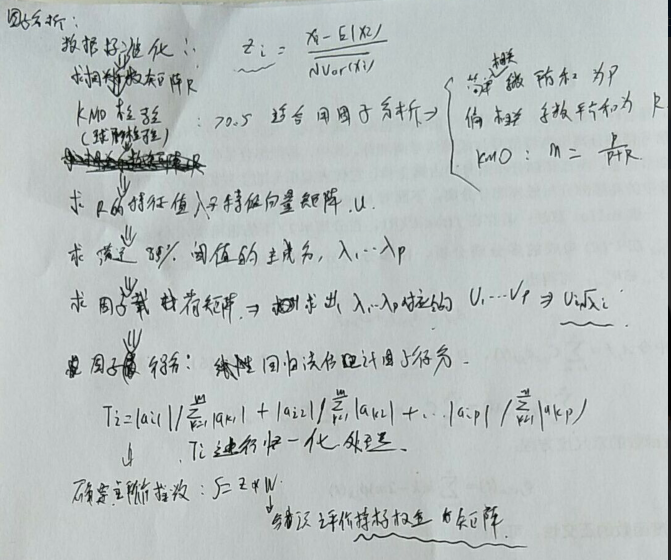


图1-3 因子分析流程图

算法步骤：

1、对原始数据进行标准化处理。为消除变量之间在数量级或量纲上的不同,需要把原始数据标准化处理。设xij表示第i(i= 1,2,… ,n)个武器第j(j= 1,2,… ,m)个指标的指标值，则xij的标准化值为：,其中，。是经过标准化变换后得到的数据矩阵：。

2、求标准化数据的相关矩阵，。

3、对相关系数矩阵进行KMO检验，当所有变量间的简单相关系数平方和远远大于偏相关系数平方和时，KMO值越接近于1,意味着变量间的相关性越强，原有变量越适合作因子分析；当所有变量间的简单相关系数平方和接近0时，KMO值越接近于0,意味着变量间的相关性越弱，原有变量越不适合作因子分析。

4、计算相关矩阵的特征值λ1≥λ2≥ … ≥λm(满足)和其对应的特征向量u1,u2,… ,um。

5、利用主成分析法计算主成分个数。找出前p个主成分,满足。

6、计算因子载荷矩阵，因子载荷 aij 的统计意义就是第i个变量与第 j 个公共因子的相关系数即表示 Xi 依赖 Fj 的份量（比重）。即表示第 i 个变量在第 j 个公共因子上的负荷，它反映了第 i 个变量在第 j 个公共因子上的相对重要性。 将前p个特征值开平方乘其对应的特称向量矩阵得到因子载荷矩阵。并对结果进行归一化处理。

7、计算指标得分（权重）：得到公因子个数p后，计算各项评价指标权重Wi，利用因子载荷矩阵求得,对进行归一化处理得到各项指标权重。

## 1.3 算法核心参考文献

中国能源安全评价的综合研究\_万佳佳

基于KMO\_Bartlett典型\_省略\_选取的PCA\_WNN短期风速预测\_解坤

计算KMO步骤：

http://www.dataguru.cn/thread-261553-1-1.html

# 2 算法开发

## 2.1算法封装（输入与输出）

|  |  |  |
| --- | --- | --- |
|  | 名称 | 属性 |
| 输入：原始数据矩阵 | data | double[][] |
| 输出：各指标权重归一化结果 | result | double[] |

## 2.2算法核心函数及说明

组成关系图参考



## 2.3 算法组成与执行流程

测试流程：

Test函数

/\*\*

\* 编写测试类

\*

\* @param data 原始数据输入矩阵

\* @param threshold 设置阈值大小（85）

\* @return 各指标权重归一化结果

\*/

流程：

数据标准化

样本相关系数矩阵计算

KMO检验

样本相关系数矩阵的特征值计算

样本相关系数矩阵的特征向量计算

获取主成分

计算因子载荷矩阵

计算各指标权重并对结果归一化

# 3 算法验证

## 3.1 验证算例说明

参考文献：中国能源安全评价的综合研究\_万佳佳

## 3.2 验证结果说明

原始数据标准化矩阵为：

-1.0638 -1.7977 1.9824 1.9915 -1.1723 -1.1434 0.0495 2.2034 -1.0158 -1.0904 2.0240 2.4275 -1.2361 0.8513

-1.0664 -1.6830 3.1406 1.3321 -1.0989 -0.2746 -0.6765 1.9108 -0.9937 -1.3878 1.7026 1.3863 -1.1760 -0.5219

-1.0999 -1.3567 1.4719 1.4784 -1.1356 -0.8131 -1.2663 1.1792 -0.9692 -1.0409 1.0761 1.5911 -1.1804 0.5145

-1.2831 -1.4572 0.7292 1.0036 -1.1968 0.7946 -1.3117 0.3848 -0.9656 -0.7930 0.5980 1.1172 -1.1816 -0.0523

-1.2934 -1.0649 -1.0434 1.0306 -1.1968 1.0449 -0.4269 0.0086 -0.9588 -1.0409 0.4002 0.6259 -1.1288 -0.8069

-1.2960 -0.4444 -0.4328 0.8033 -1.1601 2.0296 -0.1547 -0.2632 -0.9511 -0.7930 0.0292 0.6025 -1.1062 -0.3632

-0.9735 -0.6298 -0.5919 1.1836 -1.0009 -0.3650 0.2083 -0.0542 -0.9184 0.0496 -0.1521 0.3744 -1.0408 -0.5154

-0.8703 -0.5356 -0.7612 0.5941 -0.8785 -1.3932 0.8662 -0.0542 -0.8661 -0.0496 -0.1109 0.1813 -0.9120 -0.5834

-0.5814 -0.2854 -0.4393 0.2138 -0.6215 -0.6212 2.2954 0.2803 -0.7840 -0.4461 0.2683 0.1287 -0.6339 -0.1559

-0.2538 0.0395 -0.3123 -0.0045 -0.3277 0.4603 2.4089 0.6566 -0.6386 -0.3470 0.6557 0.1170 -0.3109 0.4789

0.0068 0.0356 -0.1853 -0.4253 -0.0462 1.0150 1.2972 0.7820 -0.4917 -0.4461 0.8123 0.1462 -0.0037 1.4311

0.1487 0.1868 0.2123 -0.9181 0.1741 -0.4196 0.3444 0.5939 -0.3177 -0.4461 0.6392 -0.0351 0.2296 1.5606

0.3267 0.3288 -0.1596 -1.0576 0.3945 -0.9210 0.0041 0.2803 0.0071 -0.3965 0.3342 -0.3100 0.4564 1.1687

0.4763 0.5303 0.1341 -0.6728 0.4557 0.4668 -0.6991 -0.0960 0.3009 0.0496 -0.0697 -0.5089 0.5121 0.6926

0.6775 0.7678 -0.2020 -0.9924 0.5903 0.9207 -0.2227 -0.3468 0.6195 0.0991 -0.3087 -0.6610 0.6458 0.3461

0.8555 0.8635 -0.6624 -0.6683 0.7984 0.3003 0.1856 -0.5140 0.9347 0.5452 -0.5395 -0.7721 0.8177 0.2522

1.1574 0.9508 -0.5945 -0.3668 1.0187 0.8759 0.3671 -0.6186 1.1694 0.0496 -0.5890 -0.8365 1.0954 0.3591

1.2425 0.9957 -0.4598 -0.5018 1.1411 -1.3359 -0.2681 -0.8276 1.2611 0.6939 -0.8445 -0.9359 1.1649 0.0352

1.2606 0.9962 -0.4123 -0.7786 1.2513 -1.1714 -0.3135 -1.0157 1.6195 0.9417 -1.0589 -1.0120 1.2618 -0.2045

1.2554 1.2035 -0.4662 -1.0216 1.3125 -0.7682 -0.7218 -1.2457 1.6348 1.4870 -1.3474 -1.0821 1.2638 -0.4312

1.2141 1.1537 -0.5380 -1.2241 1.3370 1.6205 -1.0621 -1.5174 1.2743 1.8835 -1.6359 -1.1523 1.2400 -0.8036

1.1600 1.2025 -0.4098 -0.9991 1.3615 -0.3019 -0.9033 -1.7264 1.0494 2.4783 -1.8832 -1.3922 1.2230 -3.2521

数据标准化完毕

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

计算样本相关系数矩阵：

1.0000 0.9431 -0.4135 -0.8870 0.9961 -0.0718 -0.1055 -0.7162 0.9656 0.8101 -0.7607 -0.8903 0.9958 -0.0971

0.9431 1.0000 -0.6426 -0.9318 0.9501 0.0670 0.0175 -0.8333 0.9025 0.8284 -0.8515 -0.9701 0.9561 -0.1504

-0.4135 -0.6426 1.0000 0.5377 -0.4320 -0.2179 -0.2812 0.7543 -0.4099 -0.5081 0.6952 0.6947 -0.4520 0.1901

-0.8870 -0.9318 0.5377 1.0000 -0.9011 -0.0891 -0.0207 0.6994 -0.8008 -0.7146 0.7089 0.9117 -0.9115 0.0010

0.9961 0.9501 -0.4320 -0.9011 1.0000 -0.0478 -0.1440 -0.7531 0.9676 0.8406 -0.7940 -0.9062 0.9982 -0.1486

-0.0718 0.0670 -0.2179 -0.0891 -0.0478 1.0000 -0.0666 -0.1647 -0.0566 -0.0539 -0.0850 -0.0782 -0.0292 -0.0283

-0.1055 0.0175 -0.2812 -0.0207 -0.1440 -0.0666 1.0000 0.2226 -0.2585 -0.2165 0.2257 0.0102 -0.1282 0.3128

-0.7162 -0.8333 0.7543 0.6994 -0.7531 -0.1647 0.2226 1.0000 -0.7829 -0.8731 0.9894 0.8921 -0.7493 0.5366

0.9656 0.9025 -0.4099 -0.8008 0.9676 -0.0566 -0.2585 -0.7829 1.0000 0.8293 -0.8201 -0.8646 0.9647 -0.2106

0.8101 0.8284 -0.5081 -0.7146 0.8406 -0.0539 -0.2165 -0.8731 0.8293 1.0000 -0.9250 -0.8401 0.8158 -0.5533

-0.7607 -0.8515 0.6952 0.7089 -0.7940 -0.0850 0.2257 0.9894 -0.8201 -0.9250 1.0000 0.9012 -0.7838 0.5568

-0.8903 -0.9701 0.6947 0.9117 -0.9062 -0.0782 0.0102 0.8921 -0.8646 -0.8401 0.9012 1.0000 -0.9108 0.2822

0.9958 0.9561 -0.4520 -0.9115 0.9982 -0.0292 -0.1282 -0.7493 0.9647 0.8158 -0.7838 -0.9108 1.0000 -0.1107

-0.0971 -0.1504 0.1901 0.0010 -0.1486 -0.0283 0.3128 0.5366 -0.2106 -0.5533 0.5568 0.2822 -0.1107 1.0000

样本相关系数矩阵计算完毕

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

KMO检验值为：0.7130

因子分析适合情况:很合适

KMO检验完毕

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

计算相关系数矩阵的特征值：

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0004 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0011 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0052 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0101 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0154 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0390 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0897 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.1525 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.3756 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.9720 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 1.3929 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 1.6494 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 9.2967

样本相关系数矩阵的特征值计算完毕

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

计算相关系数矩阵的特征向量：

0.0744 -0.6254 -0.4835 -0.2194 -0.1671 0.1954 0.1723 -0.0380 0.2059 0.1274 0.0491 -0.1919 -0.1787 -0.3074

0.0320 -0.0192 0.0369 0.3037 0.6913 0.5126 -0.1852 -0.0274 -0.0200 0.0372 0.0244 0.0534 -0.1488 -0.3193

-0.0008 -0.0401 0.0913 -0.1137 0.0334 -0.0389 -0.5337 -0.0265 -0.1049 0.5259 0.1734 -0.5732 0.0118 0.2064

0.0169 0.0562 0.1782 -0.0765 -0.1291 0.4650 -0.1037 0.0000 0.7402 -0.0562 -0.1325 -0.0198 0.2468 0.2921

0.6365 0.5026 0.0875 -0.3238 -0.0041 0.0308 0.2327 0.0016 0.0889 0.1204 0.0601 -0.1801 -0.1345 -0.3132

-0.0036 0.0114 -0.0139 0.0078 -0.0510 -0.0008 0.0130 0.0728 0.1348 0.3236 0.7844 0.4843 0.1368 -0.0143

-0.0067 0.0055 0.0273 -0.0837 -0.0026 -0.1736 -0.1584 0.0031 0.1832 0.5003 -0.4451 0.4764 -0.4754 0.0438

-0.1924 0.4078 -0.4442 0.3918 -0.1531 0.2676 0.3268 0.0712 -0.0445 0.2311 0.0316 -0.1795 -0.2483 0.2979

-0.0066 0.0364 0.0181 0.5529 -0.0111 -0.5009 -0.0410 -0.2035 0.4979 -0.0141 0.0716 -0.2221 -0.0415 -0.3070

-0.0198 -0.0718 0.1848 0.2129 -0.1852 0.0381 0.0387 0.8256 -0.0317 0.1916 -0.1603 -0.0656 0.1984 -0.2985

0.2114 -0.3869 0.5722 0.2236 0.0203 0.0526 0.4750 -0.0901 -0.0701 0.1291 0.0873 -0.1044 -0.2415 0.3048

-0.0326 -0.0409 -0.2101 -0.2365 0.6308 -0.3482 0.2573 0.3610 0.2531 -0.0096 0.0274 -0.1102 0.0507 0.3193

-0.7103 0.1634 0.3294 -0.3410 0.0286 0.0468 0.2526 -0.0819 0.0933 0.0774 0.0821 -0.1603 -0.1600 -0.3125

0.0126 0.0114 0.0668 -0.0474 -0.1298 0.0082 -0.3089 0.3444 0.1003 -0.4691 0.2923 -0.0536 -0.6579 0.1039

样本相关系数矩阵的特征向量计算完毕

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

样本相关系数矩阵的特征值矩阵(一维)为：

0.0000 0.0004 0.0011 0.0052 0.0101 0.0154 0.0390 0.0897 0.1525 0.3756 0.9720 1.3929 1.6494 9.2967

将样本相关系数矩阵的特征值矩阵中的数据从大到小排序:

9.2967 1.6494 1.3929 0.9720 0.3756 0.1525 0.0897 0.0390 0.0154 0.0101 0.0052 0.0011 0.0004 0.0000

选取前N个主成分对应的特征值数组的索引值为：

[11, 12, 13]

获取主成分完毕

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

因子载荷矩阵为：

-0.2264 -0.2295 -0.9372

0.0630 -0.1911 -0.9734

-0.6765 0.0151 0.6293

-0.0234 0.3170 0.8906

-0.2126 -0.1727 -0.9550

0.5715 0.1756 -0.0436

0.5623 -0.6106 0.1337

-0.2119 -0.3189 0.9083

-0.2622 -0.0533 -0.9361

-0.0774 0.2549 -0.9101

-0.1232 -0.3102 0.9295

-0.1301 0.0651 0.9736

-0.1892 -0.2055 -0.9529

-0.0633 -0.8449 0.3169

计算因子载荷矩阵完毕

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

各指标权重为：

0.2170 0.1621 0.2634 0.1760 0.1996 0.2193 0.3407 0.2338 0.1807 0.1773 0.2073 0.1485 0.2012 0.2733

各指标权重归一化结果为：

0.0723 0.0540 0.0878 0.0587 0.0665 0.0731 0.1136 0.0779 0.0602 0.0591 0.0691 0.0495 0.0671 0.0911

各指标权重计算完毕

流程图参考

