主成分分析算法开发文档

# 1 算法原理

## 1.1 算法概述

主成分分析就是利用“降维”的思想，在损失很少信息的前提下把多个指标转化为几个综合指标。一般来说,影响系统效能的特征参数很多,但是在建立模型时,不可能建立包括所有影响效能的参数方程体系,应该选择尽可能少的参数或参数的某种组合来建立模型。通常将多个指标化为少数指标的方法可以采用灰色关联度分析和Rough集理论中的知识约简方法来挑选关键特征参数,这两种方法实质是挑选部分指标来反映原来信息，很显然丢失了部分信息。

采用多元统计分析中的主成分分析方法来缩减指标，该方法实质是将原来的所有指标综合成尽可能少的几个综合性指标，要求这几个综合指标既能充分反映原来的指标反映的信息，又能使这几个综合指标之间互不相关。

## 1.2 算法组成与步骤

主成分分析的具体步骤:

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)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。

4)确定主成分个数。找出前p个主成分,满足。

5)前p个主成分对应的特征值组成的矩阵为Um×p,则缩减后p个综合指标为。



图2-3 主成分分析流程图

## 1.3 算法核心参考文献

<https://wenku.baidu.com/view/794757e75ef7ba0d4a733bd9.html>

基于主成分分析法的移动支付APP可用性优化设计研究

基于主成分分析的我国开放式基金风险分散优化应用研究

# 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：缩减后的综合指标矩阵 | 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

/\*\*

\* @param data 原始数据输入矩阵

\* @param threshold 设置阈值大小（85）

\* @return 缩减后的综合指标矩阵

\*/

流程：

数据标准化

样本相关系数矩阵计算

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

求取缩减后的综合指标矩阵

# 3 算法验证

## 3.1 验证算例说明

参考文献：基于主成分分析和支持向量机的效能评定\_高尚

## 3.2 验证结果说明

原始数据标准化矩阵为：

-0.7976 -1.1044 -0.9277 -0.8294 -0.5026 -0.7078 -0.7406 -0.6691 -0.4491

-0.7380 -0.7075 -0.4731 -0.2073 -0.8603 -1.0327 -0.7794 -0.8544 -1.6279

-0.7595 -0.6758 -0.7653 -0.3456 -0.8007 -0.7078 -0.7933 -0.4451 -0.8420

-0.7359 -0.6758 -0.7978 -0.4147 -0.6814 -0.7078 -0.5162 -0.6390 0.3368

0.6907 0.9831 1.3451 -0.2073 0.1533 0.9167 0.4813 0.0074 0.7298

0.6775 0.9831 1.3451 -0.2073 0.9881 0.9167 0.4813 0.6537 0.7298

1.6627 1.1973 0.2737 2.2117 1.7036 1.3228 1.8668 1.9464 1.1227

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

计算样本相关系数矩阵：

1.0000 0.9582 0.7588 0.7960 0.9589 0.9731 0.9851 0.9428 0.8028

0.9582 1.0000 0.8980 0.6362 0.8878 0.9759 0.9060 0.8427 0.7947

0.7588 0.8980 1.0000 0.2517 0.6718 0.8466 0.6467 0.5495 0.6388

0.7960 0.6362 0.2517 1.0000 0.7809 0.6475 0.8655 0.8865 0.5070

0.9589 0.8878 0.6718 0.7809 1.0000 0.9322 0.9599 0.9710 0.8027

0.9731 0.9759 0.8466 0.6475 0.9322 1.0000 0.9346 0.8840 0.8633

0.9851 0.9060 0.6467 0.8655 0.9599 0.9346 1.0000 0.9660 0.8157

0.9428 0.8427 0.5495 0.8865 0.9710 0.8840 0.9660 1.0000 0.7531

0.8028 0.7947 0.6388 0.5070 0.8027 0.8633 0.8157 0.7531 1.0000

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

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 0.0000 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 0.0148 0.0000 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0189 0.0000 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0951 0.0000 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.3337 0.0000 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.9591 0.0000

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 7.5784

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

计算相关系数矩阵的特征向量：

0.7287 -0.2404 0.2271 0.1285 0.4306 -0.0940 -0.1342 -0.0191 0.3613

-0.2583 -0.4604 0.4540 0.0193 -0.4722 -0.2580 -0.2211 0.2343 0.3485

0.1500 0.3347 -0.3972 -0.3038 -0.0939 -0.0396 -0.3688 0.6269 0.2757

0.0072 0.4831 0.0797 -0.1943 -0.1569 -0.4360 -0.2289 -0.6130 0.2821

-0.1411 0.1024 0.3515 -0.4688 0.1126 0.6916 0.0361 -0.0987 0.3524

-0.3558 0.4109 0.1460 0.6648 0.2816 0.0407 0.0228 0.1821 0.3559

-0.4197 -0.4185 -0.4810 -0.1836 0.4427 -0.1819 0.0140 -0.1639 0.3572

0.1970 -0.1371 -0.4494 0.3667 -0.4961 0.4023 0.0009 -0.2810 0.3450

0.1353 0.1078 -0.0072 -0.1521 -0.1548 -0.2410 0.8618 0.1648 0.3082

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

样本相关系数矩阵的特征值矩阵(一维)为：

0.0000 0.0000 0.0000 0.0148 0.0189 0.0951 0.3337 0.9591 7.5784

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:

7.5784 0.9591 0.3337 0.0951 0.0189 0.0148 0.0000 0.0000 0.0000

选取前N个主成分对应的特征值数组的索引值为：

[7, 8]

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

前N个主成分对应的特征值的特征向量矩阵为：

-0.0191 0.3613

0.2343 0.3485

0.6269 0.2757

-0.6130 0.2821

-0.0987 0.3524

0.1821 0.3559

-0.1639 0.3572

-0.2810 0.3450

0.1648 0.3082

缩减后的综合指标矩阵为：

-0.1605 -2.2255

-0.3248 -2.4477

-0.3453 -2.0489

-0.1322 -1.6313

1.3786 1.6841

1.1149 2.1966

-1.5306 4.4728

流程图参考

