

文章编号: 1000-8241(2014)02-0139-05

基于主成分-聚类分析法的管道风险评价方法

张杰

中国石油西部管道公司, 新疆乌鲁木齐 830012

摘要: 为了更准确地开展油气管道风险评价, 在论述主成分分析和聚类分析法基本原理的基础上, 建立了基于主成分-聚类分析法的油气管道风险评价模型, 并将此模型应用于我国某输气管道的风险评价。按照建立的风险评价流程, 通过数值标准化和统计分析, 降低了管道风险指标间的相关性和赋权主观性, 提取了管道风险指标的3个主成分因子, 然后结合指标间的相关性, 最终给出了不同管段的风险分类和排序。基于主成分-聚类分析法的管道风险评价方法能够有效提高油气管道风险评价体系的可靠性和实用性, 对于指导油气管道运营维护具有实际意义。(图2, 表6, 参6)

关键词: 油气管道; 风险评价; 主成分分析; 聚类分析

中图分类号: TE88

文献标识码: A

doi: 10.6047/j.issn.1000-8241.2014.02.005

Pipeline risk assessment method based on principle component-clustering analysis

ZHANG Jie

PetroChina West Pipeline Company, Urumqi, Xinjiang, 830012

Abstract: In order to make risk assessment on oil/gas pipeline more accurately, on the basis of principle component analysis and clustering analysis, the paper establishes risk assessment model of oil/gas pipeline based on principle component-clustering analysis, and uses this model on the risk assessment of a inland gas pipeline. According to the established risk assessment process, by using numerical standardization and statistical analysis, the correlation and weight subjectivity between pipeline risk indexes are lowered, 3 principle component factors of pipeline risk indexes are proposed, then according to the correlation between indexes, risk classification and sequence of different pipe sections are given. The pipeline risk assessment method based on principle component-clustering analysis can improve the reliability and practicality of oil/gas pipeline risk assessment system, and has practical significance for the operation and maintenance of oil/gas pipeline. (2 Figures, 6 Tables, 6 References)

Key words: oil/gas pipeline, risk assessment, principle component analysis, clustering analysis

油气管道风险评价是管道完整性管理的关键环节, 是将反映管道风险的危害因素和后果的管道属性数据、环境数据等多项指标的信息加以综合, 从总体上评价管道风险大小, 并提出相应风险控制措施的多指标综合评价、分析过程^[1-2]。多指标综合评价的基本思想是将包含各个侧面的多个单项指标组合起来形成一个综合指标, 然而多指标的综合评价难以避免评价指标之间相关性和赋权的主观性问题^[3]。

主成分分析法从数据的内部结构出发, 通过数学变化产生综合评价指标涉及的赋权, 减弱了多指标综合评价体系中的主观性影响^[3]。聚类分析是依据数据

相似度, 将大量数据点的集合分成若干类, 使得每个类中的数据间最大程度地相似, 而不同类中的数据最大程度地不同。采用主成分-聚类分析法开展管道风险评价, 在消除管道风险评价指标信息重叠和主观性的同时, 实现依据风险指标对管道管段的排序、归类, 为管道风险管理, 维抢修物资、人员的配置提供科学依据。

1 综合评价方法模型

1.1 主成分分析

主成分分析是一种数学变换方法^[3-4]。其将具有

一定相关性的一组变量 X_1, X_2, \dots, X_k , 通过线性变换, 转换为一组不相关的变量 F_1, F_2, \dots, F_k 。在这种变换中, 保持变量的总方差不变, 新变量的方差越大, 表示包含的原变量信息越多。将具有最大方差的变量 F_1 称为第一主成分, 如果第一主成分不足以代表原来 K 个指标的信息, 再考虑选取 F_2 即次大方差的变量为第二主成分, 依此类推可以构造出第三、第四、……第 P 个主成分。但在实际应用中, 为了简化, 通常不是找出 k 个主成分, 而是找出 $q (q < k)$ 个主成分即可, 条件是这 q 个主成分可以反映出原来 k 个变量的大部分方差。

主成分分析的步骤: 原始数据标准化, 建立变量的相关系数阵, 计算特征根和相应的特征向量, 确定主成分的个数, 建立主成分 (F_1, F_2, \dots, F_k) 的表达式, 建立综合指标 F 的表达式。

1.2 聚类分析^[3]

聚类分析是统计学中研究“物以类聚”问题的多元统计分析方法, 在统计分析的应用领域已经得到广泛的应用。聚类分析发展至今, 有许多种不同的聚类方法, 其中应用得最多、最成熟的方法为系统聚类法和 K -means 聚类法。在此采用系统聚类法, 其基本思想为: 首先将一定量的指标和变量各自视作一类, 然后根据指标间的亲疏程度, 合并相关性较高的类, 之后考虑合并后的类与其他类之间的亲疏程度, 再进行合并, 以使得类别内数据的差异尽可能小, 类别间数据的差异尽可能大。

2 应用实例

选取某输气管道部分管段进行风险评价研究, 管道直径 660 mm, 最小壁厚 7.1 mm, 最大操作压力为 6.3 MPa, 全线长 200 km, 埋深大部分超过 1.5 m, 管道沿线有第三方工程施工, 操作人员均经过培训。根据综合评价方法模型, 按照图 1 流程开展评价工作。

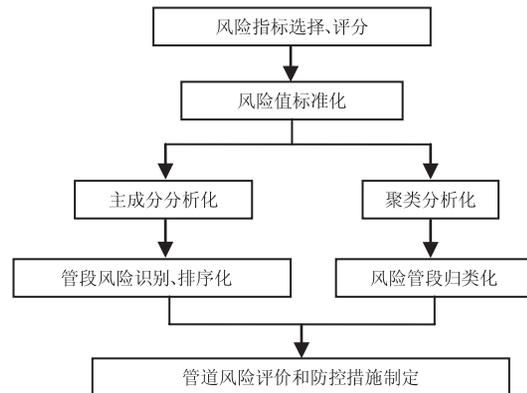


图 1 基于主成分-聚类分析法的管道风险评价流程图

2.1 管道风险因素指标选择

管道风险因素主要包括管道工程设计、施工、管道腐蚀与防腐、第三方破坏、管理和误操作 5 个方面的指标^[5-6]。在此采用基于肯特评分法^[2]对风险因素进行评分 (表 1)。按照管段实际情况选择有差异的 10 个指标: X_1 为埋深, X_2 为壁厚, X_3 为人口密度, X_4 为公众态度, X_5 为上方活动, X_6 为阴保电流, X_7 为土壤腐蚀, X_8 为杂散电流, X_9 为敷设方式, X_{10} 为土体类型。

表 1 管道风险因子量化评分结果

管段	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	1.5	7.1	280.0	424.0	4.0	430.0	90.0	112.0	459.0	453.0
2	1.5	7.1	280.0	424.0	4.0	431.0	92.0	112.0	459.0	453.0
3	1.5	8.7	280.0	424.0	4.0	431.0	92.0	112.0	459.0	453.0
4	1.5	8.7	280.0	424.0	4.0	431.0	92.0	112.0	459.0	448.0
5	1.5	8.7	280.0	424.0	4.0	431.0	90.0	112.0	455.0	453.0
6	2.0	10.3	279.0	425.0	2.0	431.0	90.0	110.0	458.0	449.0
7	2.0	10.3	277.0	425.0	2.0	431.0	92.0	112.0	458.0	449.0
8	0.8	10.3	277.0	425.0	2.0	431.0	92.0	112.0	454.0	448.0
9	2.0	10.3	277.0	425.0	2.0	431.0	92.0	112.0	454.0	448.0
10	2.0	10.3	277.0	425.0	2.0	431.0	92.0	112.0	458.0	448.0
11	2.0	10.3	277.0	425.0	2.0	431.0	92.0	377.0	456.0	448.0
12	2.0	10.3	279.0	425.0	2.0	431.0	92.0	377.0	454.0	449.0
13	2.0	10.3	279.0	425.0	2.0	431.0	90.0	377.0	456.0	446.0
14	2.0	10.3	279.0	425.0	2.0	431.0	90.0	112.0	454.0	449.0
15	2.0	10.3	279.0	425.0	4.0	431.0	92.0	112.0	458.0	449.0
16	2.0	10.3	279.0	425.0	4.0	430.0	92.0	112.0	458.0	449.0
17	2.0	10.3	274.0	425.0	4.0	430.0	92.0	112.0	458.0	449.0
18	1.6	10.3	279.0	424.0	4.0	430.0	92.0	112.0	458.0	449.0

2.2 评价指标标准化

为避免因各指标量纲的不同而缺乏可比性,首先按照下述公式对原始评价数据标准化(表 2),标记为 x_1, x_2, \dots, x_{10} 。

$$x_i = \frac{X_i - \bar{X}}{s}$$

式中: x_i 为标准化后的数据; X_i 为变量值; \bar{X} 为平均值; s 为标准差。

表 2 管道风险因子量化评分标准化数据

管段	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	-0.81	-2.31	0.96	-1.37	0.97	-1.82	-1.57	-0.43	1.05	1.70
2	-0.81	-2.31	0.96	-1.37	0.97	0.52	0.60	-0.43	1.05	1.70
3	-0.81	-0.88	0.96	-1.37	0.97	0.52	0.60	-0.43	1.05	1.70
4	-0.81	-0.88	0.96	-1.37	0.97	0.52	0.60	-0.43	1.05	-0.69
5	-0.81	-0.88	0.96	-1.37	0.97	0.52	-1.57	-0.43	-0.99	1.70
6	0.68	0.56	0.34	0.69	-0.97	0.52	-1.57	-0.45	0.54	-0.21
7	0.68	0.56	-0.89	0.69	-0.97	0.52	0.60	-0.43	0.54	-0.21
8	-2.90	0.56	-0.89	0.69	-0.97	0.52	0.60	-0.43	-1.51	-0.69
9	0.68	0.56	-0.89	0.69	-0.97	0.52	0.60	-0.43	-1.51	-0.69
10	0.68	0.56	-0.89	0.69	-0.97	0.52	0.60	-0.43	0.54	-0.69
11	0.68	0.56	-0.89	0.69	-0.97	0.52	0.60	2.17	-0.48	-0.69
12	0.68	0.56	0.34	0.69	-0.97	0.52	0.60	2.17	-1.51	-0.21
13	0.68	0.56	0.34	0.69	-0.97	0.52	-1.57	2.17	-0.48	-1.65
14	0.68	0.56	0.34	0.69	-0.97	0.52	-1.57	-0.43	-1.51	-0.21
15	0.68	0.56	0.34	0.69	0.97	0.52	0.60	-0.43	0.54	-0.21
16	0.68	0.56	0.34	0.69	0.97	-1.82	0.60	-0.43	0.54	-0.21
17	0.68	0.56	-2.75	0.69	0.97	-1.82	0.60	-0.43	0.54	-0.21
18	-0.51	0.56	0.34	-1.37	0.97	-1.82	0.60	-0.43	0.54	-0.21

2.3 主成分提取与计算

根据标准化后变量的相关系数矩阵(表 3)可知,壁厚与公众态度、土体类型,上方活动与公众态度、敷

设方式、土体类型等指标之间的相关系数绝对值较大,说明变量之间具有较强的相关性,证明其存在信息上的重叠,适合主成分分析。

表 3 相关系数矩阵

风险指标	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1.000	0.478	-0.258	0.554	-0.289	-0.005	-0.053	0.311	-0.002	-0.375
X_2	0.478	1.000	-0.568	0.811	-0.573	0.088	0.193	0.256	-0.448	-0.801
X_3	-0.258	-0.568	1.000	-0.625	0.354	0.151	-0.377	-0.032	0.194	0.477
X_4	0.554	0.811	-0.625	1.000	-0.707	0.189	0.088	0.315	-0.455	-0.715
X_5	-0.289	-0.573	0.354	-0.707	1.000	-0.535	0.124	-0.446	0.614	0.601
X_6	-0.005	0.088	0.151	0.189	-0.535	1.000	-0.033	0.238	-0.367	-0.146
X_7	-0.053	0.193	-0.377	0.088	0.124	-0.033	1.000	-0.054	0.178	-0.169
X_8	0.311	0.256	-0.032	0.315	-0.446	0.238	-0.054	1.000	-0.380	-0.390
X_9	-0.002	-0.448	0.194	-0.455	0.614	-0.367	0.178	-0.380	1.000	0.337
X_{10}	-0.375	-0.801	0.477	-0.715	0.601	-0.146	-0.169	-0.390	0.337	1.000

特征根可以被视为表示主成分影响力大小的指标,如果特征值小于 1,说明该主成分的解释力度还不如直接引入一个原变量的平均解释力度大,因此主成分个数的提取原则之一为主成分对应的特征值大于 1。根据表 4 列出的公共因子的特征根和累计方差贡献率,前 3 个主成分特征根均大于 1,并且方差累计贡献率为 72.104%(>70%),包含了原始指标大部分的信息量,将其作为反映原始指标的信息量可以认为是有效的,故提取因子个数为 3,即原来的 10 个油气管道风险评价指标可以合成 3 个公共因子。

表 4 特征根和方差贡献率数据

成分	特征根	方差贡献率%	累计贡献率%
1	4.349	43.488	43.488
2	1.774	17.738	61.226
3	1.088	10.878	72.104
4	0.841	8.409	80.513
5	0.693	6.932	87.445
6	0.465	4.647	92.092
7	0.372	3.719	95.811
8	0.169	1.686	97.497
9	0.138	1.378	98.875
10	0.113	1.125	100.000

2.4 主成分表达式的计算

表5为提取的3个主成分在原指标上的载荷和特征向量。主成分特征向量表示主成分和相应的原先变量的相关关系,可通过载荷除以主成分对应特征值的平方根计算得到,其绝对值越大,则主成分对该变量的代表性越大。依表5中可知,壁厚、公众态度、上方活动、土体类型在第一主成分上有较高载荷,说明第一主成分基本反映了这些指标的信息;阴保电流和土壤腐蚀在第二主成分上有较高载荷,说明第二主成分基本上反映了两个指标的信息;埋深在第三个主成分上有较高载荷,说明第三主成分反映了该类指标的基本信息。将得到的特征向量与标准化后的数据相乘,得出主成分表达式:

$$F_1 = 0.256x_1 + 0.419x_2 - 0.289x_3 + 0.438x_4 - 0.396x_5 + 0.151x_6 + 0.049x_7 + 0.242x_8 - 0.287x_9 - 0.401x_{10}$$

$$F_2 = 0.146x_1 + 0.188x_2 - 0.430x_3 + 0.107x_4 + 0.280x_5 - 0.487x_6 + 0.453x_7 - 0.280x_8 + 0.364x_9 - 0.119x_{10}$$

$$F_3 = 0.642x_1 + 0.005x_2 + 0.197x_3 + 0.067x_4 + 0.104x_5 - 0.362x_6 - 0.542x_7 + 0.226x_8 + 0.238x_9 + 0.006x_{10}$$

再根据特征根的大小对上面两个主成分加权,即得到下面的综合得分值:

$$F = 0.603F_1 + 0.246F_2 + 0.151F_3$$

表5 主成分在原指标上的载荷和特征向量数据

	主成分载荷			特征向量		
	第一主成分 F_1	第二主成分 F_2	第三主成分 F_3	第一主成分 F_1	第二主成分 F_2	第三主成分 F_3
X_1	0.533	0.195	0.670	0.256	0.146	0.642
X_2	0.874	0.250	0.005	0.419	0.188	0.005
X_3	-0.603	-0.573	0.206	-0.289	-0.430	0.197
X_4	0.914	0.142	0.07	0.438	0.107	0.067
X_5	-0.826	0.373	0.108	-0.396	0.280	0.104
X_6	0.315	-0.649	-0.378	0.151	-0.487	-0.362
X_7	0.103	0.604	-0.565	0.049	0.453	-0.542
X_8	0.504	-0.373	0.236	0.242	-0.280	0.226
X_9	-0.598	0.485	0.248	-0.287	0.364	0.238
X_{10}	-0.836	-0.158	0.006	-0.401	-0.119	0.006

2.5 风险管段得分及排名

表6反映了18个管段基于不同风险因素的排名。管段的排名越靠前,说明该管段受该类风险的影响越大(表6)。

表6 18个管段不同风险因素得分及排名

管段	第一主成分 F_1	排名	第二主成分 F_2	排名	第三主成分 F_3	排名	综合因子 F	排名
1	-3.876	18	-0.364	11	1.339	2	-2.225	17
2	-3.416	17	-0.520	13	-0.685	15	-2.291	18
3	-2.816	16	-0.251	9	-0.678	14	-1.862	15
4	-1.858	14	0.032	8	-0.691	16	-1.217	14
5	-2.339	15	-1.978	18	0.013	11	-1.895	16
6	0.818	9	-0.757	14	1.140	3	0.479	10
7	1.286	7	0.750	6	-0.274	12	0.919	7
8	1.151	8	-0.464	12	-3.064	18	0.118	12
9	2.066	4	0.061	7	-0.764	17	1.146	4
10	1.478	5	0.807	4	-0.277	13	1.048	5
11	2.399	1	-0.292	10	0.069	10	1.385	1
12	2.147	3	-1.254	15	0.070	9	0.997	6
13	2.321	2	-1.692	17	1.482	1	1.207	3
14	1.411	6	-1.510	16	0.657	5	0.579	8
15	0.162	11	0.764	5	0.170	7	0.311	11
16	-0.191	12	1.904	2	1.018	4	0.507	9
17	0.702	10	3.234	1	0.408	6	1.280	2
18	-1.398	13	1.511	3	0.115	8	-0.454	13

2.6 聚类分析

通过系统聚类法分析得到了各管段的聚类树状图(图2),清晰反映了各管段之间的风险因素的相似性。

图中第15~18管段聚为一类,通过分析原始指标(表1)可知,4个管段的10个评价指标中大部分指标完全一致,其余指标略有差异。

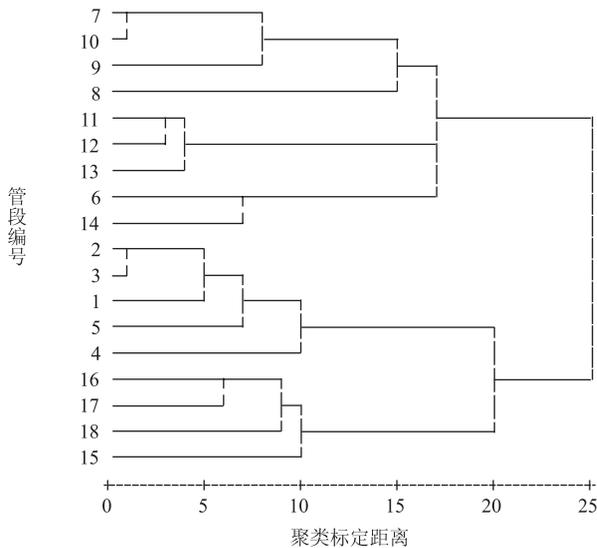


图2 不同管段聚类分析图

3 结束语

实证分析表明:主成分-聚类分析方法能够简化管道风险评价多指标,可以解决其评价系统中的主观性问题。根据各管段在不同因素中的得分及排名情况,可以分析得到各管道的主要风险因素。根据得出的不同管段风险因素的分类数据,管道企业可针对不同管段存在的风险因素,制定对应的风险控制方案和措施^[2]。讨论了主成分-聚类分析法在某管道上的应用状况,该方法也可用于并行管道、管廊带以及管道分布较集中的区域的风险评价工作,指导管道企业有针对性地进行管道维抢修力量的配置。

(上接第 138 页)

- [9] 杜伟. 低碳经济与中国石油石化行业的发展[J]. 国际石油经济, 2010(1):32-37.
- [10] Enbridge corporate. Enbridge 2012 corporate social responsibility report[EB/OL]. 2012-10-22[2013-04-10].<https://www.enbridgegas.com/assets/docs/Enbridge2012CSRReport.pdf>.
- [11] NAZARY S, GRIFFIOEN A. 管道施工中的碳足迹[EB/OL]. 2010-03-21 [2013-04-10].<http://wenku.baidu.com/view/0f1c961dc281e53a5802ff4b.html>.
- [12] 于洋曜, 潘高翔. 中国开展碳交易亟须解决的基本问题[J]. 东方法学, 2009(6):78-86.

(收稿日期:2013-04-22; 修回日期:2013-07-20; 编辑:王凯濛)

参考文献:

- [1] 李大全, 张鹏, 艾慕阳, 等. 模糊聚类法在油气管道风险评价管段划分中的应用[J]. 天然气工业, 2012, 32(7):63-67.
- [2] 戴联双, 张俊义, 张鑫, 等. RiskScore 管道风险评价方法与应用[J]. 油气储运, 2010, 29(11):818-820.
- [3] 姜扬. 聚类和主成分回归在经济指标数据中的应用研究[D]. 长春: 吉林大学, 2010.
- [4] 王慧, 王银. 主成分分析和因子分析的区别与联系及其应用[C]. 南京: 第九届中国不确定系统年会、第五届中国智能计算大会、第十三届中国青年信息与管理学者大会论文集, 2011.
- [5] 冯斌, 李大全, 廖国威. 改进的油气管道风险专家评分方法[J]. 油气储运, 2008, 27(11):4-8.
- [6] 赵忠刚, 姚安林, 赵学芬, 等. 油气管道风险因素的权重赋值方法研究[J]. 天然气工业, 2007, 27(7):103-105.

(收稿日期:2013-05-08; 修回日期:2013-08-27; 编辑:潘红丽)



作者简介: 张杰, 工程师, 1980年生, 2008年博士毕业于南京农业大学资源环境专业, 现主要从事长输油气管道的完整性管理工作。

ZHANG Jie, Ph.D, engineer, born in 1980, graduated from Nanjing Agricultural University, resources and environment, in

2008, engaged in the integrity management of long-distance oil/gas pipeline.

Tel: 0991-7806245, Email: zhangjiexbgd@petrochina.com.cn



基金项目: 中国石油天然气股份有限公司天然气与管道分公司科研项目“油气管道运营碳管理技术研究”, 20120305。

作者简介: 田望, 助理工程师, 1986年生, 2011年硕士毕业于中国科学院广州能源研究所热能工程专业, 现主要从事管道低碳、科技规划等研究工作。

TIAN Wang, MS.D, assistant engineer, born in 1986, graduated from Guangzhou Institute of Energy Conversion, CAS, thermal engineering, in 2011, engaged in the research of low carbon of pipelines and science and technology planning.

Tel: 13785674531, Email: timwang2008@126.com