

基于主成分分析与 K-L 变换的双重属性优化方法

赵加凡, 陈小宏

(石油大学 CNPC 物探重点实验室 北京 102249)

摘 要: 利用主成分分析客观赋权原理计算地震属性在预测目标参数时贡献率的大小, 通过去除权重系数较小的属性参数, 实现了地震属性的敏感性分析, 建立储层参数与有效属性之间的匹配关联; 在此基础上, 利用 K-L 变换将属性样本空间的高维属性映射为低维属性, 去除了属性之间的相关性, 有效地解决了属性组合的优化问题, 表明了主成分分析和 K-L 变换相结合的属性双重优化方法克服了单纯使用每种方法时的局限性, 充分发挥了各自的优点, 有助于属性分析、关联以及组合优化问题的解决, 提高了地震储层参数预测的运算速度和精度。

关键词: 地震数据处理; 主成分分析; K-L 变换; 神经网络; 属性组合与优化

中图分类号: N37 P631.4

文献标识码: A

文章编号: 1000-8918(2005)03-0253-04

地震属性优化是指利用专家的先验知识或数学方法, 优选出对所求解问题最敏感的、属性个数最少的地震属性或地震属性组合, 以提高储层参数预测精度, 改善与地震属性有关的处理和解释效果。属性优化是提高储层参数预测精度的重要途径^[1]。通常, 属性优化过程包括属性的敏感性分析和多属性的优化。近几年来, 围绕多属性优化问题, 涌现出很多新的方法, 但是大多数方法或只针对属性选择, 或者只针对属性样本分类, 不能系统地解决属性分析和优化问题。如何使选取的属性参数更具有代表性、如何有效地建立多属性与目标参数的关联, 以及如何快速地优化属性, 单纯地用一种方法都不可能很好地解决这个难题。

笔者提出了一种双重优化方法, 即: 利用主成分分析计算权重系数的客观赋权思想对地震属性进行敏感性分析, 建立属性与目标预测参数的关联, 然后利用 K-L (Karhunen-loeve) 变换进行属性组合的正交变换, 从高维属性中优选出低维的属性组合。因为利用主成分分析可以找出影响目标值的有效属性, 舍去权重系数特别小的、被认为不敏感的属性参数。而 K-L 变换则能将多属性样本矩阵进行维数压缩, 保留最能反映目标信息的属性与样本组合, 变换后的属性之间互不相关, 且几乎保留了所有原始数据的信息, 较合理地反映出原来对象的特征。两种方法的综合运用就可以同时解决属性和样本的优化问题, 避免了由于属性个数冗余和输入样本繁多影响计算效率和预测精度问题。

1 方法原理

1.1 主成分分析法的基本原理

主成分分析法是一种客观赋权法, 它根据各指标间的相互关系或各指标值的变异程度来确定权重系数, 能够真实地反映事物间的现实关系, 避免了人为因素带来的偏差^[2-4]。

(1) 设对某一对象进行评价时, 原始数据表中样本数为 n , 指标数为 p , 则观测样本数据矩阵为

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

实际应用中, 通常对各指标作标准化变换, 以消除原指标量纲、数量级的不同, 变换公式为

$$X_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (2)$$
$$i = 1, 2, \dots, n;$$
$$j = 1, 2, \dots, p$$

为了简化协方差矩阵的计算, 进行归一化变换:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j};$$
$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}; \quad (3)$$
$$S_j = \frac{1}{n} \sum_{i=1}^n x_{ij}^2$$

经变换后的数据均值为 0, 方差为 1。

(2) 求指标数据的相关矩阵

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & & \vdots \\ r_{n1} & \cdots & r_{np} \end{bmatrix}. \quad (4)$$

R 为对称阵,经标准化处理后的数据相关系数可以简化为

$$r_{ij} = \frac{1}{n-1} \sum_{i=1}^n x_{li} - x_{ij} \quad (5)$$

$$i, j = 1, 2, \dots, p$$

(3) 求 λ_k ($k=1, 2, \dots, p$) 的特征根和相应的特征向量 l_k ($k=1, 2, \dots, p$), 将特征根依大小顺序排列: $\lambda_1 > \lambda_2 > \dots > \lambda_p$, 其相应的特征向量记为 $l_1, l_2, \dots, l_p$, 则第 k 个主成分的方差贡献率为

$$\beta_k = \lambda_k \left(\sum_{i=1}^p \lambda_i \right)^{-1}, \quad (6)$$

前 k 个主成分的方差贡献率为 $\sum_{i=1}^k \lambda_k \left(\sum_{i=1}^p \lambda_i \right)^{-1}$ 。

(4) 选择 m 个主成分, 实际中通常取使得累计贡献率达到 85% 以上, 即 $\sum_{i=1}^m \lambda_k \left(\sum_{i=1}^p \lambda_i \right)^{-1} \geq 85\%$ 。

(5) 前 m 个主成分对总体方差的贡献矩阵 $A = (\lambda_1, \lambda_2, \dots, \lambda_m)$, 同时得到各指标在前 m 个主成分上的贡献矩阵 $L = (l_1, l_2, \dots, l_m)$, 则各指标对总体方差的贡献率矩阵 F 可由下式求出

$$F = A \cdot L = (f_1, f_2, \dots, f_p) \quad (7)$$

则 F 中各元素的值即为相应指标的权重。

1.2 K-L 变换的基本原理

对于给定信号向量 $X = [x(1), x(2), \dots, x(n)]^T$ 经 K-L 变换后变为 n 点向量 $Y = [y(1), y(2), \dots, y(n)]$, Y 的各分量之间完全去除了相关性, 且是原来所有数据 $x(i)$ ($i=1, 2, \dots, n$) 的线性组合。在用方差作为衡量标准时, K-L 变换是压缩特征空间维数的最佳标准正交变换, 该正交变换对应的向量是由样本 X 的协方差矩阵 $C_x = E\{(X - U_X)(X - U_X)^T\}$ 代表求均值计算 $U_X = E\{X\}$ 是信号 X 的均值向量) 所对应的特征向量 A 组成, 即

$$Y = A^T X = \sum_{i=1}^n x_i A_i \quad (8)$$

我们称 Y 为 X 的 K-L 变换, $X = AY = \sum_{i=1}^n x_i A_i$ 为 Y 的 K-L 展开。

K-L 变换的实现步骤: ①求取向量 X 的协方差矩阵 C_x 的特征值: $\lambda_1, \lambda_2, \dots, \lambda_n$; ②求取协方差矩阵 C_x 的特征向量: $A_1, A_2, \dots, A_n$; ③将 $A_1, A_2, \dots, A_n$ 归一化, 即 $A_i^T A_i = 1$ ($i=1, 2, \dots, n$); ④由

归一化的向量 $A_1, A_2, \dots, A_n$, 可构成归一化正交矩阵 $A = [A_1, A_2, \dots, A_n]$; ⑤由 $Y = AX$ 实现对信号 X 的 K-L 变换, 则有

$$X = A^T Y = y(1)A_1 + y(2)A_2 + \dots +$$

$$y(n)A_n = \sum_{i=1}^n y(i)A_i \quad (9)$$

如果希望对 X 做数据压缩, 那么可对 X 的变换 Y 做压缩, 即舍去 $Y = [y(1), y(2), \dots, y(n)]$ 中的一部分分量。不失一般性, 假定舍去 $y(m+1), y(m+2), \dots, y(n)$, 这样由 $y(1), y(2), \dots, y(m)$ 恢复 X 时, 将只能是对 X 的近似, 记之为 $\hat{X}$

$$\hat{X} = y(1)A_1 + y(2)A_2 + \dots +$$

$$y(n)A_n = \sum_{i=1}^n y(i)A_i \quad (10)$$

$\hat{X}$ 对 X 近似的均方根误差

$$\varepsilon = E\{[X - \hat{X}]^2\} \quad (11)$$

按照最小均方误差准则, 使用 Lagrange 条件极值法求得使 ε 小于某一指定极小值 ε_0 的基向量 A_i 及达到最小的 $\varepsilon_{\min}$ 表达式

$$C_x A_i = \lambda_i A_i, \quad (i=1, 2, \dots, n) \quad (12)$$

$$\varepsilon_{\min} = \sum_{i=m+1}^n \lambda_i \leq \varepsilon_0$$

由上式可以看出, 要使均方根误差 ε 达到最小或者满足小于指定极小值的条件, 就必须尽可能使舍去部分的基向量 A_i 对应的特征值尽可能的小, 反过来说, 若将 n 个特征值按大小顺序排列, 即 $\lambda_1 \gg \lambda_2 \gg \dots \gg \lambda_n$, 在将 $\lambda_{m+1}, \dots, \lambda_n$ 舍去后, 应使余下的 $\lambda_1, \lambda_2, \dots, \lambda_m$ 保留最大的能量, 此即 ε 最小原则。所以 K-L 变换是在最小均方误差准则意义下实现数据由高维至低维的映射, 而且能够完全去除原信号中的相关性^[5,6]。

2 应用分析

笔者采用 3 层 BP 神经网络作为预测储层目标参数, BP 神经网络用 BP 神经元的多层前向神经网络的结构形式^[7~11]。由于孔隙度的变化范围在 0 ~ 1 之间, 所以此处 BP 网络使用 Sigmoid 激活函数 $f = 1/[1 + \exp(-x)]$ 。网络的训练方法采用误差反向传播算法, 分为 2 个阶段: 第一阶段, 输入信息从输入层经隐层逐层计算出各单元的输出值; 第二阶段, 输出误差逐层向前算出隐层各单元的误差, 并用此误差修正前层权值。在实例分析中, 神经网络的输入神经元数由优化后的属性数决定, 隐层节点数参考如下公式计算^[1]

$$\lg(2K);$$
$$J = (K + I)^{1/2} + \beta;$$
$$M < \sum_{k=1}^K \alpha \frac{K}{J} \quad (13)$$
$$(K > J \text{ 时 } \alpha \frac{K}{J} = 0)。$$

其中 K 为优化后地震属性个数, M 为训练样本数 β 为 1~10 的常数。当神经网络的结构、节点性质、训练方法确定以后,网络的性质也就确定了。

笔者利用某地区的三维地震资料和测井解释资料进行孔隙度预测。根据先验知识在井旁地震道目的层段提取反映孔隙度的 12 种地震属性参数,分别为 3 种频率参数(瞬时频率、瞬时频率斜率、主频率比率) 5 种振幅参数(瞬时真振幅、瞬时正交振幅、反射强度、平均振动能量、均方根振幅) 2 种相位参数(瞬时相位余弦、瞬时相位) 2 种衰减参数(衰减度带宽、反射强度斜率)。这 12 种反应孔隙度大小的属性参数分别表示为 $\alpha(1)$ $\alpha(2)$... $\alpha(12)$ 。提取 30 个井旁地震道属性参数和测井解释孔隙度数据对作为样本,其中 15 口井的参数数据对作为 BP 神经网络训练样本,另外 15 口井的参数数据对则作为检验样本。

利用主成分分析求得 12 种地震属性参数的权重系数(表 1),该权重系数的大小反映了每种属性

参数对孔隙度的贡献率,即与孔隙度的关联程度。由表 1 可以看出主频率比率 $\alpha(3)$ 、瞬时相位余弦 $\alpha(9)$ 、瞬时相位 $\alpha(10)$ 以及衰减度带宽 $\alpha(11)$ 对孔隙度的关联很小,可以近似地将这几种属性去掉,保留其余 8 种能有效反映孔隙度信息的属性,这样就完成了属性与孔隙度的初步关联和敏感性分析。然后将这 8 种属性对应的 15 个训练样本进行 K-L 变换,经过变换之后将属性组合的样本压缩到 10 维,去除了冗余数据,实现了属性组合的优化。然后用变换后的矩阵 $Y = \{y(1), y(2), \dots, y(10)\}$ 和对应的测井孔隙度值训练神经网络,确定神经网络的性质。最后将变换矩阵乘以检验样本用训练好的 BP 网络进行预测,得到检验样本的预测值(表 2)。

为了更直观地说明双重优化的效果,分别对单纯利用 2 种方法中的每一种优化结果做了对比(表 2)。利用主成分分析和 K-L 变换优化属性的预测结果平均相对误差分别为 11.2% 和 7.9%,都大于利用双重优化方法的预测平均相对误差 4.6%。显然双重属性优化方法的预测精度高,而且优化后 BP 网络收敛的迭代次数也相对较小,加快了网络收敛速度。可以看出,主成分分析和 K-L 变换相结合的双重属性优化方法在属性分析和优化方面的效果较好,提高了 BP 神经网络的运算速度以及目标预测的精度,该方法值得进一步研究和推广应用。

表 1 影响孔隙度大小的地震属性参数权重系数

属性	$\alpha(1)$	$\alpha(2)$	$\alpha(3)$	$\alpha(4)$	$\alpha(5)$	$\alpha(6)$	$\alpha(7)$	$\alpha(8)$	$\alpha(9)$	$\alpha(10)$	$\alpha(11)$	$\alpha(12)$
权重	0.10	0.11	0.01	0.14	0.10	0.12	0.11	0.13	0.03	0.02	0.01	0.12

表 2 不同优化方法预测结果比较

检验井	测井解释孔隙度 %	灰关联优化预测孔隙度 %	相对误差 %	K-L 变换优化预测孔隙度 %	相对误差 %	双重优化预测孔隙度 %	相对误差 %
井 1	14.9	13.1	12.1	15.6	5.0	14.5	3.0
井 2	14.8	13.3	10.1	13.6	8.1	14.4	2.7
井 3	8.6	8.0	7.5	8.1	5.8	8.9	3.5
井 4	12.3	13.3	8.1	12.9	4.9	11.9	3.3
井 5	16.6	15.4	7.2	15.2	8.4	15.8	4.8
井 6	15.4	14.6	5.2	14.5	5.8	15.8	2.6
井 7	12.1	12.8	5.8	13.1	8.3	11.5	5.0
井 8	5.2	5.7	9.6	5.0	3.8	5.1	2.0
井 9	4.3	4.6	7.0	3.9	9.3	3.8	11.6
井 10	4.1	3.2	22	4.9	19.5	4.6	12.2
井 11	16.8	15.6	7.1	15.7	6.5	16.1	4.2
井 12	16.1	15.2	6.0	16.9	5.0	15.6	3.1
井 13	12.7	11.4	10.2	13.8	8.7	12.2	4.0
井 14	7.3	5.6	23.3	6.5	11.0	7.0	4.1
井 15	7.9	6.8	14.0	7.1	10.1	7.6	3.8
平均误差			11.2		7.9		4.6
迭代次数	万方数据	4200		3155		1640	

3 结 论

实际资料试算表明 ,该方法效果较好 ,主成分分析对属性有效性分析以及建立关联非常有效 ,且具有简便、所需数据量、运算量小的优点。

K-L 变换能够对较大的特征数据进行维压缩 ,进行线性变换 ,将其变成新的较小的互不相关的特征数据 ,每个新的特征数据均包含了原来所有数据的特征 ,较合理地反映出原来对象的特征。而且用 K-L 变换压缩维或去掉一些较小的特征值后造成的均方误差损失最小。

2 种方法的结合突破了单纯使用一种方法的局限性 ,较系统地解决了多属性优化问题 ,提高了属性的利用效率 ,以及目标参数预测的运算速度和预测精度。

参考文献 :

[1] 陈遵德. 储层地震属性优化方法[M]. 北京 :石油工业出版社 ,

1998.

[2] 宋之杰 ,高晓红. 一种多指标综合评价中确定指标权重的方法[J]. 燕山大学学报 2002 26(1) 20 - 22.

[3] 童其慧. 主成分分析方法在指标综合评价中的应用[J]. 北京理工大学学报(社会科学版) 2002 4(1) 59 - 61.

[4] 丁海军 ,李国诚. 油气田开发方案综合评价新方法[J]. 石油勘探与开发 ,1996 ,16(4) 29 - 32.

[5] 徐伯勋 ,白旭滨 ,于常青. 地震勘探信息技术提取、分析和预测[M]. 北京 :地质出版社 ,2001.

[6] 胡广书. 数字信号处理 :理论、算法与实现[M]. 北京 :清华大学出版社 ,1997.

[7] 文世鹏. 应用数值分析[M]. 北京 :石油工业出版社 ,1999.

[8] 许东 ,吴铮. 基于 MATLAB 6. X 的系统分析与设计——神经网络(第二版) [M]. 西安 :西安电子科技大学出版社 2002.

[9] 石广仁. 地学中的计算机应用技术[M]. 北京 :石油工业出版社 ,1999.

[10] Lippman R P. An introduction to computing with neural nets[J]. IEEE Assp Magazine ,1987 4(2) 4 - 22.

[11] Drago G P ,Ridella S. An optimum weights initialization for improving scaling relationships in BP learning[R]. Int Conf Artificial Neural Networks (Espoo ,Finland) ,June ,1991 :1519 - 1522.

DUAL OPTIMIZATION OF SEISMIC ATTRIBUTES BASED ON PRINCIPAL COMPONENT ANALYSIS AND K-L TRANSFORM

ZHAO Jia-fan ,CHEN Xiao-hong

(Key lab of Geophysical Exploration under CNPC ,University of Petroleum ,Beijing 102249 ,China)

Abstract : The deciding weight theory of Principal Components Analysis is used to compute contribution values of seismic attributes to forecasting parameters ,and the attributes sensitivity analysis is solved by getting rid of these attributes with lesser weight coefficient. In this way ,the association between reservoir parameters and attributes is established. By means of K-L transform ,higher dimensional attributes are mapped to lower dimensional ones while correlation between attributes is eliminated ,thus completing the optimization of attributes. In this paper ,the target parameters are predicted by BP neural network. The application shows that the dual optimization method combining Principal Components analysis with K-L transform overcomes the limitation of either method and at the same time possesses their respective merits. In short ,the method gives a satisfactory solution to the sensitivity analysis ,association ,and combination optimization of seismic attributes ,and eventually improves the precision ,speed ,and efficiency of reservoir prediction.

Key words : principal components analysis ,K-L transform ,neural network ,attributes combination ,attributes optimization

作者简介 :赵加凡(1973 -) ,男 ,在读博士生 ,主要从事地球物理方法与信息技术工作 ,公开发表学术论文数篇。