

# 基于 Isomap 算法的地震属性参数降维处理<sup>\*</sup>

龚灏<sup>1,2</sup> 周仲礼<sup>1,2</sup> 倪艳<sup>1</sup>

(1.成都理工大学 2.“数学地质”四川省高校重点实验室·成都理工大学)

龚灏等.基于 Isomap 算法的地震属性参数降维处理.天然气工业,2008,28(5):38-40.

**摘要** 针对非线性高维地震属性参数降维的难题,在综合分析研究工区地震属性参数数据实际特征的基础上,引入非线性降维 Isomap 算法,并基于 MATLAB 平台进行了算法的程序编制,进一步将 Isomap 降维的结果与线性的 MDS 降维结果通过小波神经网络进行检验,从算法原理的角度讨论了 Isomap 算法在地震属性参数降维处理中的可适性,表明 Isomap 具有更强的降维能力和发现数据本质结构的能力,从另一个角度提供了解决地质数据处理问题的方法。

**关键词** 地震数据处理 参数 算法分析 模糊数学

## 一、引言

长期以来,应用地震属性研究对油气储层状况进行预测一直是重要的手段和方法。由于考虑到地震属性的完整性和普适性,一般提取的地震参数比较多,因此造成了数据的高维特性。高维数据虽然提供了有关客观现象极其丰富、详细的信息,但是高维数据给随后的数据处理带来很大的困难。为了有效地应用地震属性分析技术进行预测和解释,必须结合储层物性参数对高维地震属性数据进行降维处理。

降维技术可分为线性和非线性两类。线性降维算法主要有主成分分析(PCA)、独立分量分析(ICA)、Fisher 判别分析(FDA)、多维尺度分析(MDS)等。这些方法在对高维地震属性参数进行处理时,如果参数本身是线性关系或拟线性关系时,可以得出较好的结论。非线性降维算法主要有自组织映射网络、生成拓扑映射、主曲线和表面、局部线性嵌入(LLE)、等容特征射(Isomap)等。Isomap 算法是 Tenenbaum 等人于 2000 年提出的一种新方法,在高维非线性数据处理中有较为理想的效果<sup>[1-3]</sup>。

## 二、Isomap 降维技术

### 1. Isomap 算法的基本原理<sup>[4-5]</sup>

降维算法的基本原理是针对大量高维无序的并且没有明显空间特征的数据进行分析处理,以期发

现隐藏在低维数据中有意义的低维结构。Isomap 算法结合了线性降维算法(如 PCA 和 MDS)的主要特征——计算的有效性、全局的优化性和渐近收敛性等,且比较灵活地学习到数据点的非线性结构。此算法的关键是用测地线距离代替传统的欧式距离,较好地揭示了高维数据集的内在结构。在 Isomap 算法中,对邻近点是采用欧氏距离来近似计算测地线距离,对远点则用一系列近点的欧氏距离之和来估算测地线距离。这种用测地线距离代替传统的欧氏距离的算法,可更有效地在低维空间表达高维空间的数据,较好地避免了数据信息的丢失。

### 2. Isomap 算法的步骤

#### (1) 确定在流形 $M$ 上的邻域

对于输入空间中的点  $X(i, j)$ ,其欧式距离为  $dX(i, j)$ 。将每一个点与所有的点进行比较,当两点之间的距离小于固定的半径  $\epsilon$ (或  $i$  是  $j$  的  $k$ -邻域)就认为它们是相邻的,将其连接起来,边长为  $dX(i, j)$ ,得到有权图  $G$ 。

#### (2) 估计流形 $M$ 上的测地线距离 $dM(i, j)$

初始时当  $i, j$  之间有一条边,则  $dG(i, j) = dX(i, j)$ ;否则  $dG(i, j) = \infty$ 。对所有的  $k = 1, 2, \dots, N$ ,  $dG(i, j) = \min\{dG(i, j), dG(i, k) + dG(k, j)\}$ ,这样得到矩阵  $DG = \{dG(i, j)\}$ ,它是图  $G$  中所有点间的最短路径组成的。

#### (3) 应用 CMDS 构造 $d$ 维嵌入

\* 本文受到国家自然科学基金重点项目“南方中上扬子区叠合盆地沉积充填过程与物质分布规律”(编号:40739901)的资助。

**作者简介** 龚灏,1963年生,副教授,博士;现从事储层沉积学、盆地分析、数据分析及处理等研究工作。地址:(610059)四川省成都市二仙桥东三段1号。电话:13980064155。E-mail:gonghad@cdu.edu.cn

$\tau(DG) = -\frac{1}{2} [H(DG)^2 H]$ ,  $H$ 是与  $DG$ 同阶的单位矩阵,令  $\lambda_p$ 是矩阵  $\tau(DG)$ 的第  $p$ 个特征值(特征值已按降序排列), $v_p^i$ 是第  $p$ 个特征向量的第  $i$ 个分量,令  $d$ 维嵌入向量  $y_i$ 的第  $p$ 个分量等于  $\sqrt{\lambda_p} v_p^i$ 。

Isomap 给出了确定降维维数的准则。在两种情况下 Isomap 可以确定降维维数一是当残差曲线出现拐点,二是当残差已经小到一定的阈值。

### 三、实例

以川西须家河组地区为例,研究工区砂岩储层的沉积微相是以河流沼泽沉积环境中的辫状河道砂坝、曲流河道砂坝、河口砂坝、浅湖砂坝所形成的大小不等的砂体为主。储集层的岩石类型多样。以细中粒岩屑砂岩、岩屑石英砂岩、岩屑长石砂岩为主,有少量长石砂岩、石英砂岩、钙屑砂岩等。砂粒分选中等,以次棱角状、次圆状为主,胶结物主要为方解石、白云石、自生石英、黏土等,砂岩厚度存在自西向东由厚变薄的趋势。须家河组具有砂岩丰度高异常。统计表明,上三叠统砂岩/泥岩比值为 1.88,侏罗系为 1.27,其比值与其他地区的平均值相比较,如北美是 0.69,东印度群岛是 0.24,活动带陆棚是 0.43,地台是 0.48,川西上三叠统平均要高出它们 4.7 倍。

针对研究工区的具体特点,在须四段联井地震剖面提取了过井 560、563 等地震道的振幅类、频率类等 11 个属性,这些属性是在 1 500~2 300 ms 时间内提取。根据解释成果和测井资料,选取 1 850~2 000 ms 为目标层段。对提取的地震属性用本方法做预处理。剖面属性分别构成一个  $11 \times 111$  维矩阵,11 种属性,111 个样本点。下面以井 560 地震道的数据为例来说明如何运用 Isomap 算法进行降维处理。

#### 1. 对原始数据进行预处理

由于原始数据的单位不同,数据之间的数量级差别很大,为了消除量纲的影响,对原始数据进行归一化处理即将所有数据化在 0 和 1 之间。由于数据空间较大,在此不便给出。

#### 2. 数据降维<sup>[6]</sup>

在构建近邻图时,近邻方式采用固定点数的方式  $k=6$ ,用 Floyd 算法寻找最短路径,用测地线距离矩阵代替欧氏距离矩阵构建低维空间。降维的维数根据残差的判断准则,当残差曲线出现拐点时,确定降维的维数。如图 1-a 所示,降维的维数为 3 时,残差曲线出现拐点且残差已小于 0.05,从而确定降维的结果为 3。

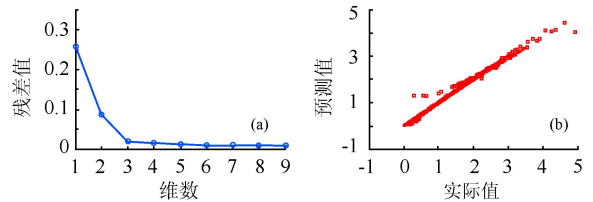


图 1 残差、拟合曲线图

再采用 MDS 方法对同一地震属性数据进行降维处理,根据 MDS 的压力函数判别法,当压力函数小于 0.05 时,确定降维的维数为 5,图 1-b 是降至 5 维时的拟合曲线图,从图中看出,降至 5 维时拟合曲线近似一条直线,效果很好<sup>[7-8]</sup>。

将 Isomap 降维以后的 3 维向量和 MDS 降维以后的 5 维向量分别作为小波神经网络的输入,这里主要采用三层的网络结构,隐层节点的个数经自适应搜索法确定,以孔隙度值为期望输出,输出层数确定为 1<sup>[9-10]</sup>,可得期望输出和预测值曲线图(图 2、3)。

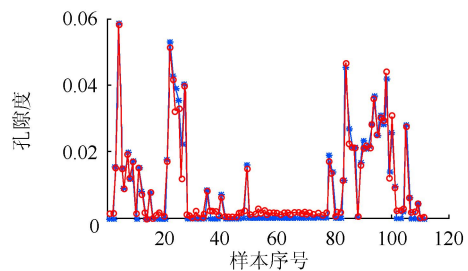


图 2 Isomap 降维预测图

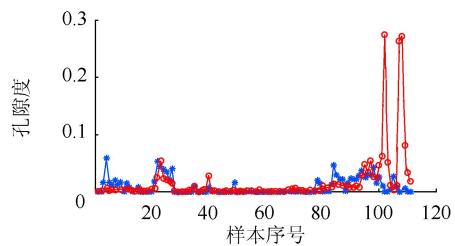


图 3 MDS 降维预测图

图中蓝点代表期望输出,红点代表预测值。从图 2 中可知 Isomap 的实际值与预测值吻合的效果比较好。图 3 是 MDS 降维后的向量作为输入,得到的预测与实际值的比较图,从图中发现有些样本的拟合误差比较大,其降维的效果远不如 Isomap 的降维效果。

### 四、结论

主要采用 Isomap 算法对须四段地震道剖面属性数据进行降维处理,在降维的过程中,近邻点  $k$  的选取起了关键的作用, $k$  的值如果过小,会使图不连

通,  $k$  值过大会使 Isomap 算法趋于 MDS 算法, 并且  $k$  的选取直接影响残差的大小。近邻点  $k$  的最优值是能够保证图连通的最小值, 笔者确定  $k$  的最优值为 6, 由图 1-a 的残差曲线出现拐点确定降维的维数, 降至 3 维时, 残差为 0.0168, 即降维后的距离矩阵与原始距离矩阵之间的相关性达到了 98.32%, 损失了极小一部分的信息。采用小波神经网络分别对 Isomap 和 MDS 两种降维的结果分别进行预测, 根据图 2 和图 3 的比较可以得出 Isomap 的预测精度明显优于 MDS 方法, 从而检验了 Isomap 降维的适用性。

将信息科学领域的 Isomap 方法引入到地质领域中, 从算法原理的角度分析了该方法是适合地质数据处理的。将该方法和 MDS 方法相比较, 说明了 Isomap 具有更强的降维能力和发现数据本质结构的能力。

实际地质数据往往含有高度的非线性, 而非线性降维方法可以发现数据集中最本质的非线性结构, 从另一个角度提供了解决地质数据处理问题的方法。

#### 参 考 文 献

- [1] JOSHUA B TENENBAUM, VIN DE SILVA, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000 (290): 2319-2323.
- [2] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290: 2323-2326.
- [3] ZHANG ZHENGYUE, ZHA HONGYUAN. Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment [R]. CSE-02-019, Technical Report, CSE, Penn State Univ, 2002.
- [4] SHIFENG WENG, CHANGSHUI ZHANG, ZHONGLIN LIN. Exploring the structure of supervised data by discriminant isometric mapping [J]. Pattern Recognition, 2005, 38: 599-601.
- [5] LEE J A, LENDASSE J A, VERLEYSEN M. Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis [J]. Neurocomputing, 2004, 57 (1): 49-76.
- [6] HONG TANG, TAO FANG, PENG-FEI SHI. Nonlinear discriminant mapping using the laplacian of a graph [J]. Pattern Recognition, 2006, 39 (1): 156-159.
- [7] 徐志节, 杨杰. 利用非线性降维方法预测膜蛋白类型 [J]. 上海交通大学学报, 2005, 39 (2): 279-281.
- [8] 赵连伟, 罗四维, 赵艳敞, 等. 高维数据流形的低维嵌入及维数研究 [J]. 软件学报, 2005, 16 (8): 1423-1430.
- [9] 侯越先, 丁峥, 何丕廉. 基于自组织的鲁棒非线性维数约减算法 [J]. 计算机研究与发展, 2005, 42 (2): 188-195.
- [10] 魏艳, 尹成, 丁峰, 等. 地震多属性综合分析的应用研究 [J]. 石油物探, 2007, 46 (1): 42-47.

(收稿日期 2008-02-14 编辑 韩晓渝)