

李海霞,吴苏怡.基于主成分分析方法的海量地震数据属性降维优化[J].地震工程学报,2019,41(3):757-762.doi:10.3969/j.issn.1000-0844.2019.03.757

LI Haixia, WU Suyi. Attribute Reduction and Optimization for Massive Seismic Data Based on Principal Component Analysis [J]. China Earthquake Engineering Journal, 2019, 41(3): 757-762. doi: 10.3969/j.issn.1000-0844.2019.03.757

基于主成分分析方法的海量地震数据属性降维优化

李海霞^{1,2}, 吴苏怡³

(1. 武汉软件工程职业学院 人文学院, 湖北 武汉 430079;

2. 华中师范大学 数学与统计学学院, 湖北 武汉 430000; 3. 东北育才学校, 辽宁 沈阳 110179)

摘要: 针对传统的地震数据属性降维优化方法所选取的地震数据属性特征贡献率低导致降维过程计算量大、CPU 占用率高等问题, 提出一种基于主成分分析的海量地震数据属性降维优化方法。首先根据地震样本特征建立地震数据特征矩阵, 把矩阵中的特征进行聚类, 运用降序法排列聚类结果, 选取前几项数据作为地震数据属性特征选取结果, 对其结果评估分类信息量; 通过特征积分准则(FSC)修正分类信息量, 获取海量地震数据属性特征节点; 运用主成分分析方法对地震数据属性特征节点主成分添加标签, 确定 Fisher 判别分析与 PCA 可变动选择不确定关系, 建立半监督降维的全局最优化形式, 运用特征值分解计算降维结果, 克服海量地震数据属性降维过程中的过拟合问题, 融合主成分分析算法与 Fisher 判别分析算法实现海量地震数据属性降维优化。实验结果证明, 所提方法选取的属性特征精度及贡献率较高, 降维过程中 CPU 占用率较低。

关键词: 地震数据特征矩阵; 降序法; PCA 算法; Fisher 判别分析算法

中图分类号: TP391.41

文献标志码: A

文章编号: 1000-0844(2019)03-0757-06

DOI: 10.3969/j.issn.1000-0844.2019.03.757

Attribute Reduction and Optimization for Massive Seismic Data Based on Principal Component Analysis

LI Haixia^{1,2}, WU Suyi³

(1. College of Humanities, Wuhan Vocational College of Software and Engineering, Wuhan 430079, Hubei, China;

2. School of Mathematics and Statistics, Huazhong Normal University, Wuhan 430000, Hubei, China;

3. Northeast Yucai School, Shenyang 110179, Liaoning, China)

Abstract: In view of the problems associated with the traditional optimization method for seismic data attribute reduction, i.e., the large amount of computation required in the reduction process and the high CPU occupancy rate, in this paper, we propose an attribute reduction and optimization method for massive seismic data based on principal component analysis (PCA). First, we establish a feature matrix of the seismic data based on the characteristics of seismic samples. The features in the matrix are then clustered and arranged in descending order. We then select the first few data as the seismic data attribute feature results and evaluate the classification information of

收稿日期: 2018-12-09

基金项目: 国家自然科学基金(4004-61772223); 武汉市教育局课题(2017118)

第一作者简介: 李海霞(1984-), 女, 湖北武汉人, 硕士研究生, 讲师, 主要研究领域为应用数学。E-mail: chenbinbin211@163.com。

these results. Next, the classification information is modified using the feature integral criterion to obtain the attribute feature nodes of the massive seismic data. We use PCA to label the principal components of the attribute nodes of the seismic data and establish a global optimization of the semi-supervised dimensionality reduction. The dimensionality reduction results are calculated by eigenvalue decomposition, we solved the problem of over-fitting in the attribute reduction process of massive seismic data, and realized the optimization of the attribute reduction of massive seismic data by combining the PCA algorithm with Fisher discriminant analysis. The experimental results show that the proposed method has a high accuracy and contribution rate of attribute feature selection, and the CPU occupation rate is low during the dimensionality reduction process.

Keywords: feature matrix of seismic data; descending method; PCA algorithm; Fisher discriminant analysis algorithm

0 引言

应用地震数据属性进行油气预测已经成为人们认识和监测油气藏的一种重要方法,而实际操作中从地震材料中获取的地震数据属性参数较为繁杂,造成了地震数据属性数据的信息冗余和浪费^[1-2],海量的地震数据属性数据虽然给我们提供了详细的信息,但会给后续的预测结果产生负面影响。在处理大量地震数据时,地震数据属性约简变得尤为重要^[3-4]。如何采用合理的地震数据属性数据准确预测含油气情况,成为当前相关研究领域亟待解决的前沿课题。在地震数据属性约简方面,相关专家学者提出了很多好的方法^[5],例如文献^[6]提出一种多目标属性降维优化方法,通过度量地震数据属性对象在近似解集上展现的冲突性,建立地震数据属性冲突信息矩阵,对矩阵元素进行特征分析,确定地震数据属性对象的重要性,实现属性特征维数约简,并采用多目标优化算法对关键地震数据属性对象集进行分解进化,最终获得属性降维优化后的近似解集。该方法选取的属性特征贡献率及累计贡献率较低。文献^[7]提出一种特征选择方法,并将其应用到地震数据属性降维优化过程中,该方法的特征选取过程是先将地震数据属性数据通过 k-近邻法拓展用于高维属性特征之间互信息的估计,再根据前向叠加策略给出全部高维属性特征最优排序结果,根据实际工作需要将冗余特征剔除,再采用后向交叉机制获取去冗后的最优强相关特征,将该特征输入到地震数据属性数据降维模型中,实现地震数据属性数据降维优化。该方法 CPU 占用率较高。文献^[8]提出一种邻域量化容差关系的地震数据属性降维优化方法,该方法首先构建地震数据属性邻域量化容差关系模型,通过该模型设定地震数据属性降维熵,

分析其相关性质,根据分析结果构造相应的地震数据属性特征选择算法,通过该算法对地震数据属性进行降维优化,该方法地震数据属性降维运行耗时较长。

针对上述问题,提出一种新的海量地震数据属性降维优化方法。构建海量地震数据的特征矩阵,并对其进行聚类处理,获取海量地震数据属性的特征节点。利用主成分分析方法对特征节点的主成分添加标签,分解海量地震数据属性的特征值,结合 Fisher 判别分析算法实现海量地震数据属性降维优化。实验仿真证明,所提方法选取的属性特征贡献率及累计贡献率较高,降维过程中 CPU 占用率较等优势。

1 基于主成分分析方法的海量地震数据属性降维优化

1.1 地震数据属性获取

首先根据地震数据属性特征构建其所属类别构成的地震数据属性描述矩阵,对地震数据属性描述矩阵中的属性表达数据进行标准化变换。其次利用属性排序方法对规范化的地震数据属性数据集进行粗选,通过特征积分准则(FSC)计算各个属性分值,最后根据属性分值降序排列结果选取排在前几的一定数量的地震数据属性作为特征选择结果。结合主成分分析方法根据特征选择结果构建属性特征主成分的自相关矩阵,对相关矩阵进行修正^[9],获取地震数据属性。

假设 $G = \{g_1, g_2, \dots, g_n\}$ 代表一个地震数据属性特征中全部属性特征组成的特征集合,其中 $g_i (1 \leq i \leq n)$ 代表任意一个地震特征, $|G| = n$ 代表所有地震数据属性个数,假设 $S = |s_1, s_2, \dots, s_m|$ 代表

由实验数据组成的地震数据属性样本集合,其中 $|S| = m$ 代表地震数据样本数量, $s_i (1 \leq i \leq m)$ 代表系统在某运行状态下所有地震数据属性的描述项,即 $s_i (1 \leq i \leq m)$ 为一个地震数据属性 n 维空间向量,且 $s_i \in R^n$ 。全部地震数据属性样本及所属类别构成属性描述矩阵为 $\mathbf{M}$:

$$\mathbf{M} = \begin{bmatrix} x_{1,1}, x_{1,2} & \cdots & x_{1,n} & l_1 \\ x_{2,1}, x_{2,2} & \cdots & x_{2,n} & l_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{m,1}, x_{m,2} & \cdots & x_{m,n} & l_m \end{bmatrix} \quad (1)$$

式中: $x_{m,n}$ 代表地震数据属性 g_m 在样本 s_n 中的属性描述项; $\mathbf{M}$ 为不同情况下的地震数据描述项。

地震数据属性描述之间存在强关联性,对于任意一个类别,会有部分属性均与类别识别关联,再对地震数据属性描述样本进行聚类、分类等数据分析之前^[10],需要对其进行预处理,为了使数据不依赖参考值,要对数据进行规范化处理,使其落在一个特定的数据区间中,将此过程称为标准化变换,具体过程如下式:

$$\bar{x}_{ik} = \frac{x_{ik} - \mu_k}{\sigma_k} \quad i=1,2,\dots,m; k=1,2,\dots,n \quad (2)$$

式中: $\bar{x}_{ik}$ 代表修正后的数据; x_{ik} 为第 i 个样本修正前的值; μ 为全部样本的平均值; σ 为全部样本的方差。

采用特征记分策略计算初始地震数据属性分值,按照降序排列方式排列地震数据属性分值,计算式如下:

$$\beta_{\text{FSC}}(g_i) = (\mu_i^+ - \mu_i^-) / (\delta_i^+ + \delta_i^-) \quad (3)$$

式中: μ_i^+ 代表地震数据属性 g_i 的正类样本的平均值; μ_i^- 代表 g_i 的负类样本的平均值; δ_i^+ 代表 g_i 的正类的标准差; δ_i^- 代表 g_i 的负类样本的标准差。地震数据属性分值的大小能够充分说明该属性可以区分不同的属性类别。为了更客观评价地震数据属性含有的分量,对 FSC 进行了修正,即:

$$\beta_{\text{FSC}}(g_i) = 0.5 |(\mu_i^+ - \mu_i^-) / (\delta_i^+ + \delta_i^-)| \quad (4)$$

依据属性 g_i 采用 Bhattacharyya 距离作为分类依据,对属性 g_i 来进行分类度量:

$$\beta_{\text{BFSC}}(g_i) = 0.25(\mu_i^+ - \mu_i^-) / [(\sigma_i^+)^2 + (\sigma_i^-)^2] \quad (5)$$

运用主成分分析方法(PCA)将高维属性数据映射至低维空间中,通过寻找一个新变量,使其反映地震数据属性的主要特征,运用 PCA 压缩地震数据属性数据空间,将高维属性数据映射至低维空间中,具体过程如下:

(1) 对规范化处理后的地震数据属性样本特征矩阵 $\bar{\mathbf{M}}$ 进行分析,计算相关系数矩阵 $\mathbf{R}$;

(2) 通过雅可比方法根据相关系数矩阵计算属性特征,特征方程 $|R - \lambda I| = 0$ 的 n 个非负特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n \geq 0, \lambda_i$ 的地震特征向量为 $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{in}) (i=1, 2, \dots, n)$, 满足:

$$\mathbf{v}_i \mathbf{v}_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases} \quad (6)$$

选择 r 个特征主要成分分量,使前面 r 个主成分

的方差占有所有方差的比例 $\eta = \sum_{i=1}^r \lambda_i / \sum_{j=1}^n \lambda_j$ 接近 1,并使选取的这 r 个主成分较多的保留初始 n 个属性的信息,得到主成分矩阵 $\bar{\mathbf{M}}_r, \bar{\mathbf{M}}_r$ 的样本类别信息仍保持。

(3) 计算 r 个属性特征主成分因子:

$$\mathbf{Y}_i = v_{i1}x_{i,1} + v_{i2}x_{i,2} + \dots + v_{in}x_{i,n} \quad (7)$$

假设 $\mathbf{R}$ 是对称的半正定矩阵,设 $\mathbf{R} = \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{bmatrix}$, $\mathbf{M}$ 全部不小于 0 的特征值的个数等于 $\mathbf{M}$ 的秩,其中 $\mathbf{A}$ 与 $\mathbf{B}$ 为相关矩阵的对角阵,分别代表地震特征主成分信息的自相关矩阵,从另一个角度看, $\mathbf{M}_{m \times n} = \mathbf{M}_{m \times (n_1+n_2)}$, 即:

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,n_1} \\ \vdots & & \vdots \\ x_{m,1} & \cdots & x_{m,n_1} \end{bmatrix} \begin{bmatrix} x_{1,n_1+1} & \cdots & x_{1,n_1+n_2} \\ \vdots & & \vdots \\ x_{m,m_1+1} & \cdots & x_{m,n_1+n_2} \end{bmatrix} \quad (8)$$

$\mathbf{A}$ 是 n_1 个地震数据属性间的相关矩阵, $\mathbf{B}$ 为 n_2 个地震数据属性间的相关矩阵, $\mathbf{C}$ 为 n_1 个地震数据属性与 n_2 个地震数据属性间的相关矩阵,由上述过程可以给出: $\mathbf{R}$ 中地震数据属性特征值大于 0 的个数为 n_r , $\mathbf{M}$ 和 $\mathbf{B}$ 中地震数据属性特征值大于 0 的个数为 n_a 与 n_b , 则 $n_r \leq n_a + \max\{n_a, n_b\}$ 。

对上述相关矩阵总的主成分进行修正,具体过程如下式:

$$\mathbf{Y}_{i1} = v'_{i1}x_{i,1} + \dots + v'_{i1}x_{i,n_1} \quad (9)$$

式中: $\mathbf{v}$ 为传统方法计算得到的地震特征向量; $\mathbf{v}'$ 为所提方法得到的地震特征向量。

根据公式(9)中获得的 $\mathbf{v}'$ 进行地震数据属性选择:

$$\mathbf{v}'(\mathbf{w}^T x_i + w_0) \geq 1 \quad i=1, 2, \dots, N \quad (10)$$

式中: $\mathbf{w}$ 为 d 维向量; w_0 为偏移量。

1.2 海量地震数据属性降维优化

对 1.1 节选择的地震数据属性添加标签,建立 Fisher 判别分析和主成分分析可变动选取的不确定关系,构建地震数据属性降维的最优化形式,通过地

震数据属性特征值分解计算降维结果,克服地震数据属性描述数据降维过程中存在的过拟合情况,实现地震数据属性降维优化,具体过程如下所述:

地震数据属性近邻数据的一个重要假设是:与目标属性主成分近邻的数据最有可能和其属于相同类别,于是设定 $\mathbf{H}$ 为 $n \times n$ 矩阵, $\mathbf{H}$ 可以压制任何与目标主成分近邻的数据在其潜在低维空间分离。常用的近邻数据处理方法是与目标主成分最近的数据设定在热核上,每一对最近邻数据根据这两者之间的关联性做出相应的惩罚,关联性依据输入的地震数据属性主成分中的数据对之间的欧式距离创建,即 $\mathbf{H}_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$, 其中, σ 表示变量,该变量可以控制矩阵 $\mathbf{H}$ 的扩展范围和幅度。进一步假设 σ_i 表示与该属性主成分 x_i 邻近的其他主成分之间的相关程度, k 表示 x_i 最近邻的主成分数量,此时有:

$$\sigma_i = \|x_i - x_i^k\|, O_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j}\right) \quad (11)$$

式中: O_{ij} 表示 x_i 与 x_j 的距离远近关系,取值在区间 $[0, 1]$ 内。假设, x_i 与 x_j 之间的距离较大, O_{ij} 的值将非常接近 0; 相反 x_i 与 x_j 之间的距离较小时, O_{ij} 的值与 1 较为接近。降维过程中 O_{ij} 处理的数据对属于相同类别的标签数据,当采用 O_{ij} 抑制随机选取标签数据的影响时, Fisher 判别分析的最优化公式中 P^b 与 P^w 可利用数据对权值思想描述,假设满足 $y_i = y_j$, 则有:

$$\mathbf{Q}_{ij}^b = \left(\frac{1}{n'} - \frac{1}{n'_{yi}}\right) A_{ij}, \mathbf{Q}_{ij}^w = O_{ij} / n'_{yi} \quad (12)$$

否则满足 $y_i \neq y_j$, $\mathbf{Q}_{ij}^b = \frac{1}{n'}$, $\mathbf{Q}_{ij}^w = 0$ 。

为了使用标签数据的同时也能够采用无标签数据,实现地震数据属性降维,对变量给出以下要求:

$$\begin{cases} P^{rb} = (1 - \delta)P^b + \delta P^r \\ P^{rw} = (1 - \delta)P^w + \delta I_d \end{cases} \quad (13)$$

式中: δ 表示变量,区间在 $[0, 1]$ 内,求解结果侧重于在 Fisher 判别分析和主成分分析之间变化。在地震数据属性主成分成对表达的基础上,找到一个转换矩阵 $\mathbf{W}$, 该矩阵满足在地震数据属性主成分样本潜在空间内类间距尽可能最小,同时使类间距离为最大,即:

$$\mathbf{W} = \arg \max_{\mathbf{W} \in R^d} [\text{trace}(\mathbf{W}^T P^{rb} \mathbf{W} (\mathbf{W}^T P^{rw} \mathbf{W}))] \quad (14)$$

从上式中可以看出,若采用 $\{\lambda_k\}_{k=1}^d$ 描述通过优化过程求解获得的地震数据属性特征值, $\{\varphi_k\}_{k=1}^d$

表示与特征值相应的向量,地震数据属性降维优化问题可通过转置矩阵获得:

$$\mathbf{W} = (\lambda_1 \varphi_1 \mid \cdots \mid \lambda_1 \varphi_r) \quad (15)$$

针对地震数据属性无监督降维过程无法利用标签信息,仅仅是在特征向量数量为给定的前提下,实现对原始地震数据属性数据潜在空间最大近似。监督降维过分拟合于标签数据,利用可调参数 δ 的半监督降维,降低了地震数据属性监督或无监督降维过程的不足,同时变量 δ 有效弥补了标签数据选择的随机性。原始地震数据属性主成分样本数量共有 n 个,其中含有 n' 个带标签的地震数据属性数据,剩余 $n - n'$ 个是无标签的地震数据属性数据,假设 y_i 是在样本 x_i 中的标签数据,可表示成形式 $\{(x_i, y_i)\}_{i=1}^{n'}$, 且满足 $y_i \in \{1, 2, \dots, c\}$, c 表示地震数据属性主成分信息类别数量, d 表示原始地震数据属性主成分的维数, r 表示降维处理后的维数。输入: 带有标签的地震数据属性数据为 $\{(x_i, y_i) \mid x_i \in R^d\}_{i=1}^{n'}$, 无标签地震数据属性主成分数据为 $\{x_i \mid x_i \in R^d\}_{i=1}^{n'}$ 。

输出: $d \times r$ 的转置矩阵 $\mathbf{W}$ 。

(1) 对于 $1 \leq i \leq n'$, 在 $\{x_j\}_{j=1}^n$ 中获取 x_i 的最近邻数据,采用 A_{ij} 求解 x_j ;

(2) 假设满足 $y_i = y_j$, 则有 $\mathbf{Q}_{ij}^b = \left(\frac{1}{n'} - \frac{1}{n'_{yi}}\right) A_{ij}$,

$\mathbf{Q}_{ij}^w = \frac{A_{ij}}{n'_{yi}}$, 否则 $y_i \neq y_j$, 则有 $\mathbf{Q}_{ij}^b = \frac{1}{n'}$, $\mathbf{Q}_{ij}^w = 0$;

(3) 求出 P^b 和 P^w ;

(4) 用特征值与特征向量的关联性,求解 $\mathbf{W}$ 。

上述过程中, A_{ij} 可以压制地震数据属性数据任意近邻的属性数据在其所属低维空间分离,处理相同类别的属性数据中的标签数据,有效避免了标签数据随机选取盲目性带来的影响,有利于后续地震数据属性降维分析;系数 δ 可以将 Fisher 判别分析算法与主成分分析算法有效融合,根据广义特征值的概念求解地震数据属性数据的主要成分生成的矩阵,获取地震数据属性数据潜在低维空间,避免监督降维过程过分拟合于标签数据。针对主成分分析算法未考虑到不同类别的属性信息的不足,利用了 Fisher 判别分析算法的优势,通过调整 δ 使地震数据属性数据降维的损失达到最小;符号 δ 为标量,区间在 $[0, 1]$ 内,该值的变化可满足 $P^{rb} \varphi = \lambda P^{rw} \varphi$, 当满足 $\delta = 0$ 时,公式转换成 Fisher 判别分析;当满足 $\delta = 1$ 时,公式转换成主成分分析。

2 仿真实验证明

为了验证所提基于主成分分析的海量地震数据属性降维优化方法的综合有效性,需要进行一次仿真实验,仿真实验环境为 Dell PC。处理器为 Intel Core-M480I5CPU@2.67 GHz、4 GB 内存,操作系统为 32 位 Windows7 系统,安装并使用了 VC++6.0、Cyg Win、Matlab R2011b 和 NS2-allinone2.26 软件。

选取文献[6]方法与文献[7]方法作为实验对比方法,对比不同方法提取的地震数据属性特征精度(%)、贡献度(%),结果如表 1、表 2、表 3 所列。表 1 中, N 表示海量地震波数据量,单位为条,用 g 表示; P 代表海量地震数据提取精度,单位为(%); R 代表贡献度,单位为(%);采用 4 种不同地震类型(S)数据:构造地震、火山地震、塌陷地震、诱发地震表示。其中,贡献度为可提取地震数据的属性特征数量与总属性特征数量的比值。

表 1 文献[6]方法属性降维优化对比结果

Table 1 Comparison results of attribute reduction and optimization using the method in reference[6]

S	N/g	$P/\%$	$R/\%$
构造地震	100	78	82
火山地震	100	73	89
塌陷地震	100	71	89
诱发地震	100	63	96

表 2 文献[7]方法属性降维优化对比结果

Table 2 Comparison results of attribute reduction and optimization using the method in reference[7]

S	N/g	$P/\%$	$R/\%$
构造地震	100	88	83
火山地震	100	79	75
塌陷地震	100	81	73
诱发地震	100	73	96

表 3 所提方法属性降维优化对比结果

Table 3 Comparison results of attribute reduction and optimization using the proposed method

S	N/g	$P/\%$	$R/\%$
构造地震	100	98	92
火山地震	100	93	91
塌陷地震	100	92	90
诱发地震	100	99	94

分析表 1、表 2、表 3 可知,在不同数据量下,所提方法的属性降维效果明显优于文献[6]方法与文献[7]方法,当构造地震类型的数据量为 100 个时,所提方法选取的属性特征精度为 98%,文献[6]方法与文献[7]方法均低于所提方法。综合 4 种数据

类型来看,所提方法选取的属性特征精度、贡献率均在 90%以上,侧面反映出所提方法更适合对大量地震数据属性数据进行降维。而文献[6]方法选取的属性特征贡献率相比其他两项高,主要原因在于该方法需要建立地震数据属性信息矩阵对数据进行检验,此过程过于繁琐,面对大量地震数据属性数据降维时,选取的属性特征贡献率较低,使得最终的降维效果不好。

对比所提方法和文献[6]与文献[7]方法地震数据属性降维过程中的 CPU 占用率(%),对比结果如图 1 所示。

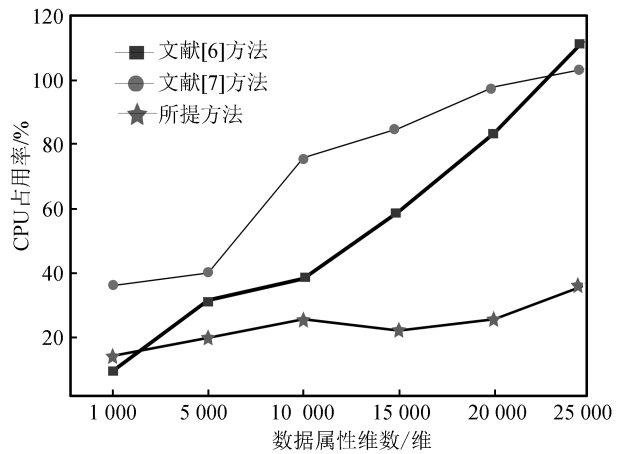


图 1 不同方法属性降维 CPU 占用率对比
Fig.1 Comparison between CPU occupancy rates for different method

由图 1 可知,所提方法与文献[6]方法、文献[7]方法相比,CPU 占用率较低,当属性维数为 1 000 时,文献[7]方法的 CPU 占用率比所提方法 CPU 占用率低 2%,而当属性维度在 10 000 时,所提方法的 CPU 占用率为 21%,而文献[7]方法的占用率达到 30%。由上述实验数据可知,随着属性维数的不断增加,不同方法进行属性降维优化,CPU 占用率也持续缓慢增长,当属性维度在 10 000 以上时,文献[7]方法进行属性降维 CPU 占用率持续上升,侧面反映出为文献[7]方法进行属性降维过程中 CPU 占用情况具有不稳定性,与此同时也反映出所提方法明显优于传统方法,更具有使用价值。

3 结论

为提高海量地震数据属性特征的选取精度,提出了一种新的低 CPU 占用率的海量地震数据属性降维优化方法。建立海量地震数据特征矩阵,聚类

处理对其进行排列选取,评估其分类信息量。通过寻找一个新变量,获取海量地震数据属性特征节点,反映地震数据属性的主要特征。运用主成分分析方法(PCA)压缩地震数据属性数据空间,将高维属性数据映射至低维空间中,实现海量地震数据属性降维优化。实验结果表明,所提方法与传统的方法相比,选取的属性特征精度及贡献率较高,说明所选取的属性特征精确描述了属性数据所含有的信息,将其用于后续属性降维过程中,可减少计算量,使得降维过程 CPU 占用率较低,具有一定的实际应用价值。

鉴于论文篇幅所限,完成时间有限,有关详细的技术环节这里没有能做更多详述,有兴趣的读者可与作者联系讨论。作者水平有限,文中遗漏之处及错误难免,望请读者给予批评指正。

参考文献(References)

- [1] 张浩.网络数据特征选择的优化方法研究与仿真[J].计算机仿真,2017,34(2):367-370.
ZHANG Hao. Network Data Feature Selection Research and Simulation Optimization Method [J]. Computer Simulation, 2017, 34(2): 367-370.
- [2] 陈翔,陆凌姣,吉人,等.SBFS:基于搜索的软件缺陷预测特征选择框架[J].计算机应用研究,2017,34(4):1105-1108,1119.
CHEN Xiang, LU Lingjiao, JI Ren, et al. SBFS: Search Based Feature Selection Framework for Software Defect Prediction [J]. Application Research of Computers, 2017, 34(4): 1105-1108, 1119.
- [3] 丁强,李锴,朱泽磊,等.基于变量降维的大电网经济调度优化方法[J].电力系统自动化,2017,41(18):130-136.
DING Qiang, LI Kai, ZHU Zelei, et al. Economic Dispatch Optimization Method for Large-Scale Power Grid Based on Variable Dimension Reduction [J]. Automation of Electric Power Systems, 2017, 41(18): 130-136.
- [4] 李万益,孙季丰.基于高斯增量降维与流形 Boltzmann 优化的人体运动形态估计[J].电子学报,2017,45(12):3060-3069.

- LI Wanyi, SUN Jifeng. Human Motion Estimation Based on Gaussian Incremental Dimension Reduction and Manifold Boltzmann Optimization [J]. Acta Electronica Sinica, 2017, 45(12): 3060-3069.
- [5] 黄南天,彭华,蔡国伟,等.电能质量复合扰动特征选择与最优决策树构建[J].中国电机工程学报,2017,37(3):776-785.
HUANG Nantian, PENG Hua, CAI Guowei, et al. Feature Selection and Optimal Decision Tree Construction of Complex Power Quality Disturbances [J]. Proceedings of the CSEE, 2017, 37(3): 776-785.
- [6] 罗乃丽,李霞,王娜.利用冲突信息降维的进化高维目标优化算法[J].信号处理,2017,33(9):1169-1178.
LUO Naili, LI Xia, WANG Na. Objective Reduction Using Conflict Information in Many-Objective Optimization Evolutionary Algorithm [J]. Journal of Signal Processing, 2017, 33(9): 1169-1178.
- [7] 周红标,乔俊飞.基于高维 k-近邻互信息的特征选择方法[J].智能系统学报,2017,12(5):595-600.
ZHOU Hongbiao, QIAO Junfei. Feature Selection Method Based on High Dimensional K-Nearest Neighbors Mutual Information [J]. CAAI Transactions on Intelligent Systems, 2017, 12(5): 595-600.
- [8] 姚晟,徐风,赵鹏,等.基于邻域量化容差关系粗糙集模型的特征选择算法[J].模式识别与人工智能,2017,30(5):416-428.
YAO Sheng, XU Feng, ZHAO Peng, et al. Feature Selection Algorithm Based on Neighborhood Valued Tolerance Relation Rough Set Model [J]. Pattern Recognition and Artificial Intelligence, 2017, 30(5): 416-428.
- [9] 祝璞,黄章进.基于稀疏贝叶斯模型的特征选择[J].计算机工程,2017,43(4):183-187,193.
ZHU Pu, HUANG Zhangjin. Feature Selection Based on Sparse Bayesian Model [J]. Computer Engineering, 2017, 43(4): 183-187, 193.
- [10] 刘航.考虑不确定性的地震灾害应急救援设备并行优化设计[J].地震工程学报,2018,40(5):1118-1123.
LIU Hang. Parallel Optimization Design, Incorporating Uncertainty, of Emergency Rescue Equipment for Earthquake Disasters [J]. China Earthquake Engineering Journal, 2018, 40(5): 1118-1123.