

# 石油勘探电磁层析成像优化检测研究

敬明旻, 费海涛

(中国石化石油工程技术研究院信息与标准化研究所, 北京 100101)

**摘要:**在对石油勘探电磁层析成像数据的高效检测过程中,由于成像过程受到电磁回波的干扰,成像信号容易衰减,导致层析成像效果差,后期检测过程需要消耗大量的时间进行数据比对,图像挖掘准确率低的问题。提出改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方法,先将电磁层析成像的时序数据进行离散傅里叶变换,使其从时域空间转换至频域空间,并将电磁层析成像的时序数据映射到多维空间的点上,搜寻出每个电磁层析成像时序数据点的相似性,融合聚类算法对石油勘探电磁层析成像数据集进行聚类,通过计算异常电磁层析成像数据集的局部偏离因子,实现了改进聚类算法的石油勘探电磁层析成像数据的高效检测方案。仿真结果证明,基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方法精确度高,鲁棒性强,为石油勘探优化提供了科学依据。

**关键词:**时间序列;数据挖掘;聚类算法

**中图分类号:**TP311 **文献标识码:**B

## Petroleum Exploration and Electromagnetic Tomography Optimizing Mining Detection Research

JING Ming-min, FEI Hai-tao

(Research Department of Information and Standardization, Sinopec Research Institute of Petroleum Engineering,  
Beijing 100101, China)

**ABSTRACT:** For oil exploration in high efficiency mining of electromagnetic tomography data detection process, the imaging process by the interference of electromagnetic echo imaging signal attenuation, easily lead to poor tomographic imaging effect, later need to consume a lot of time for data mining process, the problem of low accuracy of image mining. Based on improved clustering algorithm of petroleum exploration data of electromagnetic tomography efficient mining method. This method divides the time-series data of electromagnetic tomography discrete Fourier transform, make the space transform from time domain to frequency domain space, and time sequence of electromagnetic tomography data mapping to the multidimensional space point, search out every time sequence of electromagnetic tomography data points of similarity, fusion clustering algorithm of petroleum exploration data sets of electromagnetic tomography clustering, abnormal electromagnetic tomography data set by calculation of partial deviation factor, realized based on the improved clustering algorithm of petroleum exploration data of electromagnetic tomography efficient mining scheme. Simulation experiment proves that the improved clustering algorithm based petroleum exploration data of electromagnetic tomography efficient mining method is of high precision and strong robustness.

**KEYWORDS:** Time series; Data mining; Clustering algorithm

### 1 引言

近些年,随着开采技术水平的不断提高,电磁层析成像技术因其具有低成本、非侵入的优势,在勘探领域占据着主导地位<sup>[1-3]</sup>。在对石油岩层进行勘探时,是通过电磁波产生的图像来对石油进行勘探的<sup>[4-6]</sup>。由于石油区域的岩层结构较为错综复杂,当电磁层析成像处于分布不均的岩层时,

电磁波的回波过程容易出现混淆,存在电磁层析图像产生局部易变的隐患<sup>[7]</sup>。而石油勘探电磁层析成像数据的高效挖掘方法是消除这一隐患的有效途径,并引起了很多专家与学者的重视。由于石油勘探电磁层析成像数据的高效挖掘方法具有广阔的发展空间,因此,也成为了业内人士研究的焦点课题,受到了广泛的关注,也出现了很多好的方法。

文献[8]提出了利用蚂蚁的强大搜索能力来实现石油勘探电磁层析成像数据的高效挖掘。该算法适应性强,但是存在容易陷入局部最优解的问题。文献[9]提出了通过对石油

收稿日期:2014-12-04 修回日期:2015-04-17

勘探电磁层析成像数据的分析,处理数据并构建模糊型电磁层析成像数据模型实现对石油勘探电磁层析成像数据的高效挖掘的方法。该方法精确度较高,但是存在耗时长,适应性差的弊端。文献[10]中重点研究了基于关联规则算法的石油勘探电磁层析成像数据的高效挖掘。该方法通过挖掘石油勘探电磁层析成像数据集,完成对石油勘探电磁层析成像数据的高效挖掘。该方法具有耗时短的优势。可是在进行层析成像数据挖掘时,电磁回波时的信号容易衰减,导致存在层析成像效果差,挖掘准确率低的问题<sup>[11]</sup>。

针对上述问题的产生,提出了基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方法。该方法先将电磁层析成像的时序数据进行离散傅里叶变换,使其从时域空间转换至频域空间,并将电磁层析成像时序数据映射到多维空间的点上,搜寻每个电磁层析成像时序数据点的相似性,融合聚类算法对石油勘探电磁层析成像数据集进行聚类,通过计算异常电磁层析成像数据集的局部偏离因子,实现了基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方案。仿真结果证明,基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方法精确度高,鲁棒性强。

## 2 石油勘探电磁层析成像数据采集原理

在石油勘探电磁层析成像数据的高效挖掘过程中,电磁层析成像数据挖掘带有强大的驱动特性,可以依据石油勘探电磁层析成像模型中含有的有效信息集合,组建其电磁层析成像数据的挖掘环境,紧密的对勘探挖掘的电磁层析成像数据实行驱动类分析,体现出石油勘探电磁层析成像数据挖掘技术的层次性,满足石油勘探电磁层析成像数据分析对精确度的要求。具体的步骤如下详述:

在石油勘探电磁层析成像数据的高效挖掘的过程中,依据电磁感应理论和高斯定论,设定电磁层析的 Maxwell 方程组的微分形式,用下式进行表述

$$\begin{cases} \nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t} \\ \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \\ \nabla \cdot \vec{B} = 0 \\ \nabla \cdot \vec{D} = \rho \end{cases} \quad (1)$$

在上式中,由 $\nabla$ 代表石油勘探电磁层析成像数据挖掘过程中的拉普拉斯算子,且必须满足于下式

$$\nabla = \frac{\partial}{\partial x} \vec{i} + \frac{\partial}{\partial y} \vec{j} + \frac{\partial}{\partial z} \vec{k} \quad (2)$$

在式(1)中,由 $\vec{H}$ 代表电磁层析成像数据挖掘中的磁场强度,由 $\vec{J}$ 代表电磁层析成像数据挖掘中的电流密度。

组建石油勘探电磁层析成像数据的挖掘环境,对被测岩层区域进行假设,并结合石油勘探的物场媒质特性,得到电磁层析成像信号概率密度函数 $\vec{A}$ 的方程为

$$\nabla(\nu^{-1} \nabla \vec{A}) = j\omega\sigma\vec{A} \quad (3)$$

式中,由 $\nu$ 代表电磁层析成像数据挖掘过程中的相对磁导率,由 $\sigma$ 代表电磁层析成像数据挖掘过程中的电导率。

依据电磁场感应理论,可以进一步说明,在石油勘探电磁层析成像数据的挖掘过程中,由于电场会随着磁场的变化而变化,因此,石油勘探电磁层析成像数据模型可以表述为

$$U = -\frac{d\varphi}{dt} = -n \frac{d(\vec{B}S)}{dt} = -n_j \omega l (\vec{A}_1 - \vec{A}_2) \quad (4)$$

综上所述,可以说明石油勘探电磁层析成像数据的高效挖掘原理,但采用当前算法进行成像数据挖掘时,电磁回波时的信号容易衰减,导致存在层析成像效果差,挖掘准确率低的问题。

## 3 电磁层析成像数据的高效挖掘优化方法

采用当前方法进行层析成像数据挖掘时,电磁回波时的信号容易衰减,导致存在层析成像效果差,挖掘准确率低的问题。为此,提出改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方法。

### 3.1 电磁层析成像时序数据傅立叶变换

在石油勘探电磁层析成像数据的优化挖掘过程中,电磁层析成像的时序数据会根据挖掘精度的要求,将其分类为系列子集,并采用离散傅立叶变换将这些电磁层析成像系列子集的时序数据从时域空间转换至频域空间,依据 Parseval 理论,使电磁层析成像的时域能量函数和频域能量函数相同,而且电磁层析成像的时域空间的所有能量聚集在前几个系数上,这样不仅降低了电磁层析成像的时序数据维数,还缩小了计算的复杂性。具体的步骤如下详述:

在石油勘探电磁层析成像数据的优化挖掘过程中,假设给定一个由 $\vec{X} = \{x_i, i = 0, 1, 2, \dots, N-1\}$ 代表的电磁层析成像的时间序列,将由 $\vec{X}$ 代表的电磁层析成像的时间序列分类为由 $\vec{X}_1, \dots, \vec{X}_m$ 代表的子序列,且满足于 $\vec{X}_1 = \{x_{ij}, j = 0, 1, 2, \dots, n-1\}, N = nm$ 的条件。对由 $\vec{X}_i$ 代表的 $\vec{X}$ 的子序列完成离散傅立叶变换可以得到下式

$$x_f = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} x_i \exp(-i * 2\pi f i / n) \quad (5)$$

在石油勘探电磁层析成像数据的优化挖掘过程中,根据 Parseval 的理论,依据下式的计算使其电磁层析成像的时域能量函数和频域能量函数相同

$$\|\vec{X}_i - \vec{Y}_i\|^2 = \|\vec{X}_f - \vec{Y}_f\|^2 \quad (6)$$

在石油勘探电磁层析成像数据的优化挖掘过程中,对反映在由 $k$ 代表维空间内的层析成像的时序点,用由 $D^k(P)$ 作为标准来衡量层析成像数据的相似度。具体的步骤如下详述:

步骤1:在石油勘探电磁层析成像数据的优化挖掘过程中,设有 $N$ 个石油勘探电磁层析成像数据点的集合数 $k$ ,随机取 $P \in k$ ,将 $D(P, x)$ 当做由 $p$ 代表的电磁层析成像点和由 $x \in p$ 代表的电磁层析成像点之间的距离,在 $k$ 的集合中,对于不包含由 $p$ 代表的石油勘探电磁层析成像数据点集合的 $D(P, x)$ ,按次序排列,得到其石油勘探电磁层析成像序列,

用下式代表

$$(D(P, x_1), D(P, x_2), \dots, D(P, x_{N-1})) \quad (7)$$

步骤2:在石油勘探电磁层析成像数据的优化挖掘过程中,对于电磁层析成像集合 $k$ 中的全部点,给定由 $n, k$ 代表的整数,可以定义为由 $p \in k$ 代表的维空间内电磁层析成像点的 $D^k(P)$ 与由 $r$ 代表维空间内电磁层析成像参考点的 $D^k(r)$ 方差,用下式能够计算得到:

$$Diff = |pD^k(P) - D^k(r)| \quad (8)$$

以 $D^k(P)$ 为标准来衡量与由 $r$ 代表的维空间内电磁层析成像参考点的 $D^k(r)$ 。

在石油勘探电磁层析成像数据的优化挖掘过程中, $D^k(P)$ 与由 $r$ 代表的维空间内电磁层析成像参考点的 $D^k(r)$ 越接近,那么,就可以认定石油勘探电磁层析成像数据点 $p$ 相近范围内的成像数据与参考点 $r$ 电磁层析成像数据越相似。

综上所述,可以说明在优化过程中,先将电磁层析成像的时序数据进行离散傅里叶变换,再使其从时域空间转换至频域空间,并将电磁层析成像时序数据映射到多维空间的点上,搜寻出每个电磁层析成像时序数据点的相似性,为实现基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘提供了有利的条件。

### 3.2 电磁层析成像数据的高效挖掘实现

在石油勘探电磁层析成像数据的优化挖掘过程中,搜寻出每个电磁层析成像时序数据点的相似性,融合聚类算法对石油勘探电磁层析成像数据集进行聚类,形成部分聚类簇和异常电磁层析成像数据集,通过计算聚类簇的质心和簇中每个电磁层析成像时序数据与质心间距的平均值与异常电磁层析成像数据集的局部偏离因子,实现了基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘。具体的步骤如下详述:

1)在石油勘探电磁层析成像数据的优化挖掘过程中,需要计算电磁层析成像 $p$ 的 $k$ 距离。

在优化挖掘中,相对于正数 $k$ ,用 $k - dis(P)$ 表述电磁层析成像数据 $p$ 到 $k$ 间的距离,而 $k - dis(P)$ 需要满足于下述2个条件才能够与 $d[p, o]$ 相吻合。而 $d[p, o]$ 代表了石油勘探电磁层析成像数据集中,由 $p$ 电磁层析成像数据和由 $o$ 代表的电磁层析成像数据间的欧式距离。

①在石油勘探电磁层析成像数据的优化挖掘过程中,石油勘探电磁层析成像数据集内最少要有 $k$ 个电磁层析成像数据,且 $o' \in D/\{P\}$ 满足 $d[p, o'] \leq d[p, o]$ 的条件成立。

②在石油勘探电磁层析成像数据的优化挖掘过程中,最多不能超过 $k - 1$ 个电磁层析成像数据,且 $o' \in D/\{P\}$ 满足 $d[p, o'] < d[p, o]$ 的条件必须成立。

2)在石油勘探电磁层析成像数据的优化挖掘过程中,电磁层析成像数据 $p$ 到 $k$ 的距离邻域计算。

由 $N_{k-dis(P)}$ 代表电磁层析成像数据 $p$ 的最近 $k$ 个邻居集合。它是由与 $p$ 间的距离小于或等于 $k - dis(P)$ 的电磁层析成像全部数据的组成由 $q$ 代表,通过下式可以表述:

$$N_{k-dis(P)} = \{q \in D/\{P\} \mid d(p, q) \leq k - dis(P)\} \quad (9)$$

3)在石油勘探电磁层析成像数据的优化挖掘过程中,电

磁层析成像数据 $p$ 的局部可达密度,可以通过下式表述

$$lrd(P) = 1 / \left[ \frac{\sum_{q \in N_k(P)} reach - disk}{|N_k(P)|} \right] \quad (10)$$

4)在石油勘探电磁层析成像数据的优化挖掘过程中,对于给定的正数需要满足于下式

$$LOF_k(P) = \frac{\sum_{q \in N_k(P)} \frac{lrd(q)}{lrd(P)}}{|N_k(P)|} \quad (11)$$

式中,可以将由 $p$ 代表电磁层析成像数据的局部偏离因子定义为由 $p$ 代表的电磁层析成像数据的异常程度,假如石油勘探电磁层析成像数据局部偏离因子越大,则可以判断为电磁层析成像数据异常的可能性就会加大,反之,则是异常的可能性越小。

综上所述可以说明,改进方法优于传统方法,其精确度高,实用性强。

## 4 实验分析

为了证明改进方法的有效性,需要进行一次实验。实验的数据采用 shuttle 数据集,此项数据集中含有 42500 个数据对象。其中采用 1000 个石油勘探电磁层析成像数据作为初始勘探电磁层析成像数据集,新增加的电磁层析成像数据个数分别表述为 2, 20, 60, 120, 320, 520。

实验 1)

分别采用传统算法与改进算法进行石油勘探电磁层析成像数据的高效挖掘。对其挖掘时间进行对比,结果如表 1 所示。从表 1 中可以看出,当新增的石油勘探电磁层析成像数据个数较少时,改进算法的时间就已经明显的优于传统算法,当新增的石油勘探电磁层析成像数据个数达到 300 以上时,改进算法的运行时间的效率更是明显高于传统算法。

表 1 不同算法运行时间比较

新增数据对象个数/(个)	传统算法 (ms)	改进算法 (ms)
2	1	0.2
20	1.2	0.4
60	1.4	0.6
120	1.45	0.8
320	1.6	0.9
520	1.95	1

实验 2)

在 shuttle 数据的实验数据集中,大部分的石油勘探电磁层析成像数据都会标记为正常数据,异常的石油勘探电磁层析成像数据是极少数的,所以如果不是石油勘探电磁层析成像的正常数据,那么,就可以定义为异常的石油勘探电磁层析成像数据,将 1000 个石油勘探电磁层析成像数据作为初始石油勘探电磁层析成像数据集,取 1000 个石油勘探电磁层析成像数据作为新增加的石油勘探电磁层析成像数据,在

初始与新增加的石油勘探电磁层析成像数据集中,不是正常的电磁层析成像数据的总数为 30,40,50,60。

采用传统算法与改进算法进行石油勘探电磁层析成像数据的高效挖掘实验,将挖掘的精确进行比较,比较的结果见表 2。

表 2 不同算法精度比较

非类 1 数据个数	正确发现的异常数据个数	
	传统算法(个)	改进算法(个)
30	10	12
40	14	19
50	17	28
60	24	39

从表 2 可以说明,改进算法的挖掘精确较比传统算法具有明显的优势,那是因为改进算法通过对电磁层析成像时序数据进行离散傅立叶变换,使其从时域空间变换到频域空间,并将电磁层析成像时序数据映射到多维空间的点,融合聚类算法对电磁层析成像数据集进行聚类,形成一些聚类簇和异常电磁层析成像数据集,对于异常电磁层析成像数据集中的数据对象,计算其局部偏离因子,证明了改进方法的高效挖掘,保障了改进算法的精确度。

实验 3)

取 100 个石油勘探电磁层析成像数据当做初始数据,将不同的算法进行石油勘探电磁层析成像数据的挖掘耗能量进行对比,对比结果见图 1。

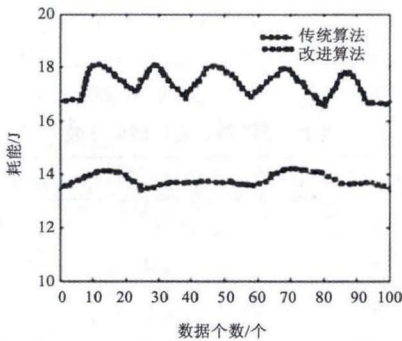


图 1 不同算法的耗能性能对比

从图 1 中可以看出改进算法不仅在挖掘精确性方面优于传统算法,在耗能方面的性能上更突出明显的优势,能够满足其石油勘探电磁层析成像数据高效挖掘的应用需求。

上述实验可以证明,改进方法的挖掘精确度高,鲁棒性强。

5 结束语

采用当前方法进行层析成像数据挖掘时,电磁回波时的信号容易衰减,导致存在层析成像效果差,挖掘准确率低

问题,提出了改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方法。该方法先将电磁层析成像的时序数据进行离散傅里叶变换,再使其从时域空间转换至频域空间,并将电磁层析成像时序数据映射到多维空间的点上,然后搜寻出每个电磁层析成像时序数据点的相似性,融合聚类算法对石油勘探电磁层析成像数据集进行聚类,通过计算异常电磁层析成像数据集的局部偏离因子,实现了基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘。实验仿真证明,基于改进聚类算法的石油勘探电磁层析成像数据的高效挖掘方法精确度高,鲁棒性强。

参考文献:

[1] 李振春. 把握行业发展方向,引领科技创新潮流——2012 年《石油地球物理勘探》评述[J]. 石油地球物理勘探, 2013-5: 824-842.

[2] 李慧,成德安,金婧. 网格层析成像速度建模方法与应用[J]. 石油地球物理勘探, 2013. A01:12-16.

[3] 李庆春,叶佩. 初至波与反射波旅行时多尺度渐进联合层析成像[J]. 石油地球物理勘探, 2013-4:536-544.

[4] 张凯,李振春,秦宁,杨晓东,关键,张敏. 非常规海相碳酸盐岩油气藏地震层析研究进展[J]. 地球物理学进展, 2013-6: 2993-3000.

[5] 徐海,樊太亮,赵夏,翟立新,张翊孟,杨剑,韩瑞冬. 基于矢量中值滤波的 ISS 高密度地震数据去噪方法[J]. 石油地球物理勘探, 2013-3:340-344.

[6] 刘勇锋,马海涛. 地震波真反射层析成像超前预报布置参数优化[J]. 中国安全生产科学技术, 2013-2:29-34.

[7] 秦宁,李振春,杨晓东,岳玉波,张凯,张晴. 高斯波束角道集的旅行时层析速度分析[J]. 石油地球物理勘探, 2013-3:366-371.

[8] 王宝华,吴淑红,韩大匡,桓冠仁,李巧云,李小波,李华,周久宁. 大规模油藏数值模拟的块压缩存储及求解[J]. 石油勘探与开发, 2013-4:462-467.

[9] 文华,余年,郭小平,雷旭友,程凯. 基于 DSSA 的电磁波 CT 软件开发及其在岩溶勘探中的应用[J]. 高速铁路技术, 2014-1: 21-26.

[10] 王敬,杨艳,胡英,张小磊,张东,秦前清. 网格尺度对走时层析成像结果的影响分析及对策[J]. 武汉大学学报(理学版), 2013-4:398-402.

[11] 张志美,等. 石油勘探电磁层析成像在复杂岩层的应用仿真[J]. 计算机仿真, 2014,31(9):407-410.

[作者简介]



敬明旻(1984-),男(汉族),四川广安人,硕士研究生,主要研究方向:电子信息;

费海涛(1967-),男(汉族),辽宁昌图人,硕士研究生,主要研究方向:计算机科学。