

学校编码: 10384

分类号___密级___

学号: 24320141152395

UDC_____

厦 门 大 学

硕 士 学 位 论 文

基于地震数据的储层预测自动寻优模型研究

Research on Reservoir Prediction Automatic Optimization
Model Based on Seismic Data

鲍彬彬

指导教师: 吴清强 副教授

专业名称: 计算机科学与技术

论文提交日期: 2017 年 4 月

论文答辩日期: 2017 年 5 月

学位授予日期: 2017 年 6 月

指导教师: _____

答辩委员会主席: _____

2017 年 4 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):
年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

地震数据中蕴含着丰富的有关地下的地质构造、底层岩性和流体物性等信息，是储层预测的数据基础。但随着我国油气勘探开发的水平越来越高和勘探领域的不断延伸，地震属性数量和种类膨胀式增长，储层预测难度增大。数据挖掘是一种有效进行储层预测手段。本文基于数据挖掘中的特征降维方法和聚类分析构造了储层预测模型。

基于数据挖掘的储层预测模型在实际使用中经常需要人工设置参数。如果使用者不了解数据挖掘技术，设置参数是一个困难的事。储层预测模型参数选择的好坏对其预测结果有很大影响。储层预测模型一般会有多个参数，有的参数变化范围比较大，模型参数取值组合很多，人工寻找模型最佳参数很耗费时间。本文提出一个储层预测模型参数自动寻优的方法，无需用户设置模型参数，减轻了用户的负担，并且收敛速度快。

首先，针对储层预测模型中的聚类算法参数优化问题，本文使用一种基于混合编码的粒子群对聚类算法的参数进行寻优。根据聚类目标—类内相似度最大、类间相似度最小，设计粒子适应度函数，进而提高聚类结果质量。其次，对整个储层预测模型的参数优化进行研究。储层预测模型包含特征降维和聚类分析两部分。通常这两部分是单独进行优化，但这样得到的结果是局部最优，而不是全局最优。本文利用混合粒子群对整个模型的所有参数同时寻优，也就是特征降维和聚类分析联合优化。最后，地震数据绝大部分都是无标签数据，标签数量很少。本文提出了一种基于聚类结果的自动标签选取方法以获得更多的标签，使得依赖于一定规模的训练数据的有监督分类算法能产生较为精确的分类器。对于基于分类算法的储层预测模型参数寻优，也使用混合粒子群方法，以提高分类精确度。实验表明，混合粒子群能自动地对储层预测模型参数寻优，提高模型预测精确度。储层预测整个模型参数同时寻优，比单独寻优的效果好。基于自动标签的分类效果并不比聚类效果差，说明了基于聚类结果的自动标签选取方法是有效的。

关键词：储层预测模型；粒子群；自动寻优

Abstract

The seismic data contains abundant information of stratigraphic lithology, fluid physical properties and geological structure. The information about reservoir property and lithology which is effectively extracted from seismic data has very strong practicability. However, with the development of oil-gas exploration development in our country and the constant extension of exploration fields. The number and kinds of seismic attributes increase rapidly. It makes reservoir prediction much more difficult. Data mining is an effective method for reservoir prediction. In this dissertation, the reservoir prediction model is constructed based on feature dimension reduction and clustering analysis.

The reservoir prediction model based on data mining often needs some parameters to be set manually. If a user does not understand data mining technology, parameters setting is very difficult for him. Reservoir prediction model parameters Selection has great impact on its predicted results. Reservoir prediction model. Some of them vary over a wide range. There are many combination of model parameters values. It takes time to find the optimal parameters of reservoir prediction manually. This dissertation presents a method of automatically optimizing reservoir prediction model parameters. The model parameters input is not required. The method reduces the burden of users. It can converge to global optimal solution rapidly.

Firstly, aiming at the problem of clustering algorithm parameters optimization in reservoir prediction model, in this dissertation, a hybrid particle swarm algorithm is presented to be used to optimize the clustering algorithm parameters. According to clustering objective - the maximum inner-class similarity and the minimum between-class similarity, the fitness function is designed to improve the quality of clustering results. Secondly, the parameters optimization of the reservoir prediction model is studied. The reservoir prediction model includes two parts: feature reduction and cluster analysis. These two parts are usually optimized separately. But the obtained results are local optimal, and not overall optimal. This dissertation applies hybrid particle swarm algorithm to optimize all parameters of the model simultaneously, which jointly optimizes feature dimension reduction and clustering analysis. Finally, the vast majority of seismic data is unlabeled data, the number of seismic labels is small. Therefore, this dissertation proposes a method of label data selection based on clustering results to obtain more seismic labels. It makes supervised classification

algorithms which depends on a certain number of training data can generate more accurate classifiers. This dissertation uses the hybrid particle swarm algorithm to optimize the parameters of reservoir prediction model based classification algorithms, which can improve classification accuracy. Experiments indicate that Hybrid particle swarm optimization method can automatically optimize the parameters of reservoir prediction model. The accuracy of model prediction has been improved. Optimizing all the parameters of Reservoir prediction model at the same time is better than optimizing them separately. Classification results based on automatic labels is not worse than clustering results. This shows that the automatic labels selection method based on clustering results is effective.

Key words: Reservoir Prediction Model; Particle Swarm Optimization; Automatic Optimization

目 录

第一章 绪论	1
1.1 本文的研究背景及意义	1
1.2 参数寻优研究现状	2
1.3 本文的研究内容和组织结构	4
第二章 储层预测模型及其参数寻优研究现状	6
2.1 储层预测模型介绍	6
2.1.1 聚类	7
2.1.2 特征选择	7
2.1.3 特征抽取	8
2.2 特征降维的参数寻优研究现状	9
2.3 聚类的参数寻优研究现状	11
2.4 主要问题和研究路线	11
2.5 本章小结	13
第三章 聚类算法参数寻优	14
3.1 智能优化算法	14
3.2 粒子群优化算法	14
3.2.1 基本粒子群优化算法	14
3.2.2 二进制粒子群优化算法	16
3.3 基于混合粒子群算法的聚类参数寻优	17
3.3.1 混合编码	17
3.3.2 适应度函数设计	17
3.4 实验分析	19
3.4.1 实验数据	19
3.4.2 聚类有效性评价指标	20
3.4.3 实验结果	22

3.5 本章小结.....	26
第四章 储层预测模型参数寻优.....	27
4.1 特征降维的评价.....	27
4.2 基于混合粒子群的储层预测模型参数寻优.....	27
4.3 实验分析.....	31
4.4 本章小结.....	41
第五章 基于分类算法的储层模型参数寻优.....	42
5.1 基于类几何结构和类概率的标签数据选取方法.....	42
5.2 分类算法在地震数据中的应用.....	43
5.3 基于分类和混合粒子群的储层预测模型参数寻优.....	45
5.4 实验分析.....	47
5.5 本章小结.....	49
第六章 总结与展望.....	51
6.1 工作总结.....	51
6.2 展望.....	52
参考文献.....	53
附录.....	57
致谢.....	58

Contents

Chapter 1 Introduction	1
1.1 Research Background and Significance.....	1
1.2 Research Status of Parameter Optimization.....	2
1.3 Research Content and Organizational Structure	4
Chapter 2 Reservoir Prediction Model and Its Parameter Optimization Research Progress	6
2.1 Introduction of Reservoir Prediction Model	6
2.1.1 Clustering.....	7
2.1.2 Feature Selection.....	7
2.1.3 Feature Extraction.....	8
2.2 Research Status of Optimizing Feature Dimension Reduction Parameters.....	9
2.3 Research Status of Optimizing Clustering Parameters.....	11
2.4 Main Problem and Research Route	11
2.5 Summary.....	13
Chapter 3 Clustering Algorithms Parameters Optimization	14
3.1 Intelligent Optimization Algorithms	14
3.2 Particle Swarm Optimization	14
3.2.1 Basic Particle Swarm Optimization	14
3.2.2 Binary Particle Swarm Optimization	16
3.3 Clustering Algorithms Parameters Optimization Based on Hybrid Particle Swarm	17
3.3.1 Hybrid Coding	17
3.3.2 Fitness Function.....	17
3.4 Experiment Analysis.....	19
3.4.1 Experimental Data	19
3.4.2 Evaluation Indexes of Clustering Validity.....	20
3.4.3 Experimental Result.....	22
3.5 Summary.....	26

Chapter 4 Reservoir Prediction Model Parameters Optimization ...	27
4.1 Evaluation of Feature Dimension Reduction	27
4.2 Reservoir Prediction Model Parameters Optimization Based on Hybrid Particle Swarm	27
4.3 Experiment Analysis	31
4.4 Summary	41
Chapter 5 Reservoir Prediction Model Parameters Optimization Based on Classification	42
5.1 Label Data Selection Method Based on Class Geometrical Structure and Class Probability	42
5.2 Application of Classification in Seismic Data	43
5.3 Reservoir Prediction Model Parameters Optimization Based on Classification and Hybrid Particle Swarm	45
5.4 Experiment Analysis	47
5.5 Summary	49
Chapter 6 Conclusions	51
6.1 Conclusions	51
6.2 Future Work	52
References	53
Appendix	57
Acknowledgements	58

第一章 绪论

1.1 本文的研究背景及意义

随着世界经济的飞快发展，各国对能源的需求不断增大，石油作为能源领域里最为重要的战略资源，早已成为影响国家经济政治发展至关重要的因素。石油是工业生产和航天军工最主要的能源和原料，其化工产品在日常生活中广泛使用。没有石油充分的供给，无法保障国家经济快速稳定地发展，国防安全也会受到影响^[1-2]。

储层是指具有连通孔隙的岩层，能在其中储集和渗滤油气^[3]。地震勘探作为石油勘探的重要手段之一，是利用地下岩层介质弹性不同，通过研究人工方法激发的地震波在地下不同岩层内的传播规律与特点，推断出地下的地质情况的地球物理勘探方法^[4]。地震数据是地震波经过转化得到的。地震属性是地震数据基于地震几何学，地震动力学，地震波运动学等理论使用相关计算方法得出的数据，蕴含着地下岩性、流体物性和含油气性等信息，可以在不同程度上反映地层地质的信息^[5]。在油气勘探的第一阶段中，主要是确定一块新区域是否存在油气，但在该阶段钻井相对较少，造成测井资料不足，难以推断地下岩层地质情况。而含有丰富地下地质信息的地震属性数据能够弥补该阶段测井资料的不足^[6]。基于地震属性分析技术的储层预测在实际的勘探开发中得到了广泛的应用，并获得了良好效果。

我国近年来一直不断提高油气田勘探开发的程度，不断扩张勘探领域。地震属性数量和种类随之膨胀式增长，储层预测难度增大。数据挖掘是一种有效进行储层预测手段。数据挖掘是从大量的，有噪声的实际应用数据中发现隐含在其中有用的信息和知识的过程^[7]。面对地震属性多而杂的问题，可以利用数据挖掘中特征选择方法从众多的地震属性中选取对储层特征敏感的属性或者使用特征抽取方法从众多的地震属性中提取最能反映储层特征的新属性。油井造价昂贵，现今绝大多数地震属性数据都是无标签数据，可以利用聚类分析对地震属性数据进行分类，并结合测井数据分析地质情况。聚类分析是一种无监督数据挖掘方法，不需要标签信息的指导，自主将样本数据划分为若干个类。其划分目标是同一个类别中数据尽可能地相似，不同类别中的数据尽可能地不同

[8]。聚类分析在语音识别、系统建模、文本挖掘、图像处理等领域有着广泛的应用^[9]。

基于数据挖掘构建预测模型时，该模型经常是带有参数的。如果不了解数据挖掘技术，设置参数是一个困难的事。预测模型应用中，模型参数的质量对其聚类结果有很大影响。为获取最优结果，一个最简单的方法是手动调整模型参数，一个个去试。这样的方法枯燥，工作量大。本文对预测模型自动化进行了研究，不需要人工设置参数，并自动寻找最优参数。本文具有一定的研究价值。

1.2 参数寻优研究现状

针对如何进行参数寻优问题，很多学者提出了自己的解决方案。Joachims 等人^[10]提出留一法误差方法，用于比较不同参数下的支持向量机分类结果质量，选择最优参数。该方法比交叉验证法更精确，在计算方面很有优势。2002 年，Chapelle、Bousquet 等人^[11]率先提出了支持向量机参数寻优概念，并使用梯度下降法对支持向量机参数进行寻优。李琼和董才林等人^[12]在 2010 年提出了一种对支持向量机的核参数快速寻优的方法，这种方法不需要训练支持向量机分类器，通过计算各类别的可分性直接确定最优参数，大大缩短了时间，并且分类精度没有损失。

随着研究的深入，研究学者们尝试利用智能优化算法自动优化参数，智能优化算法是信息科学、生命科学与认知科学等多个学科相互交叉形成的产物。在仿生和拟物的思想指导下，智能优化算法基于对于生物智能机制和某些社会、自然规律的研究，通过数据运算的方法模拟实现生物智能机制和社会、自然规律等。智能优化具有自适应性，自组织性，通用性，不依赖于问题本身的特点。智能优化算法一般采用群体协作的方式对问题进行优化求解，适合大规模并行处理，并且具有很好的鲁棒性，对初始条件不敏感，能在不同条坏下寻找最优解。学者们对智能优化的研究在算法理论和算法性能这两方面都取了很大的进展，而且已经被广泛应用于科学研究和生产实践的各种领域中。常用的智能优化算法有遗传算法（Genetic Algorithms, GA）、粒子群算法（Particle Swarm Optimization, PSO）、模拟退火算法（Simulated Annealing, SA）、差分进化算法（Differential Evolution, DE）、蚁群算法（Ant Colony Optimization, ACO）和

人工鱼群算法 (Artificial Fish-Swarm Algorithm, AFSA) 等。2014 年, Li X Z 等人^[13]提出了一种基于遗传算法的 SVM 参数寻优方法来提高滑坡位移速率预测精确度。Sherin B M 等人^[14]于 2015 年对基于 SVM 的水下目标分类器使用遗传算法对 SVM 参数、核和核参数进行优化选择提高分类器的性能。2006 年, 邵信光、杨慧中等人^[15]提出使用基于粒子群算法的 SVM 参数优化方法建立聚丙烯腈生产过程中数均分子量的软测量模型, 以 k-fold 交叉验证最小化为寻优目标, 搜索核参数和惩罚参数的最优值。Katherasan D 等人^[16]于 2014 年使用人工神经网络 (ANN) 模拟药芯焊丝电弧焊过程中的焊缝几何形状, 并使用粒子群算法对工艺参数进行寻优, 而且根据研究领域知识, 提出用最大化的渗透深度、最小的胎圈宽度和最小化的加固作为寻优目标。2017 年, Zhang M 等人^[17]提出了一种基于支持向量机和粒子群优化算法的抑郁症识别方法。这方法利用粒子群算法对支持向量机参数进行寻优, 并使用反馈变异粒子群算法来平衡局部搜索和全局搜索, 避免粒子群优化算法陷入局部最优, 提高了支持向量机分类器的精度。Sartakhti J S 等人^[18]在 2012 年提出了一种支持向量机和模拟退火算法结合的肝炎疾病诊断模型, 通过模拟退火算法搜索支持向量机核参数的最优值。2014 年, 尤丽华和吴静静等人^[19]为了克服传统 BP 神经网络的连接权值容易陷入局部最优的缺点, 利用模拟退火算法优化 BP 神经网络, 提高了水质 PH 值预测准确度。改进的 BP 神经网络利用模拟退火算法优化训练样本的选择和神经网络结构, 提高了 BP 神经网络的泛化能力。与非线性回归预测对比, 改进的 BP 神经网络对水质 PH 预测的拟合效果更好, 精确度更高。2009 年, Qin A K 等人^[20]提出一种自适应差分进化算法, 根据以往的经验自适应调整试验个体选择策略及其相关控制参数。BP 神经网络算法容易受连接权值的初始值和阈值的影响, 导致陷入局部最优。为了解决该问题, 王林和彭璐等人^[21]使用自适应差分进化算法中的自适应变异和交叉因子对 BP 神经网络算法中的连接权值初始值和阈值进行寻优。然后用寻优得到的最优连接权值初始值和阈值训练 BP 神经网络。与一般的神经网络、自回归积分滑动平均预测模型 (ARIMA) 及其它混合预测模型对比 实验结果表明基于自适应差分进化的 BP 神经网络的预测精度更高。2010 年, Zhang X L 等人^[22]提出使用蚁群优化算法解决支持向量机的参数寻优问题, 并使用该方法对文本进行分类, 经实验表明可以产生良好的分类结果。Zhou H 等人^[23]于 2012 年提出一种基于支持向量回归的氮氧化合物排放模型,

并使用蚁群算法对范化参数和高斯核参数进行寻优，与 BP 神经网络和广义回归神经网络对比，该方法预测精确度高。2011 年，Shen W 等人^[24]提出了一种人工鱼群参数寻优的径向基函数神经网络，用于预测上海股票交易所的股票指数。Wang G P 等人^[25]在 2015 年使用支持向量回归预测城市水资源、电力、燃气等短期资源的短期需求，并通过人工鱼群算法对支持向量回归参数进行寻优，与 BP 神经网络和基于网格法参数寻优的支持向量回归对比，该方法预测更准确。

然而这些方法中没有一种是适用于任何数据挖掘方法组合成的储层预测模型。

1.3 本文的研究内容和组织结构

本文主要提出了一种储层预测自动化寻优模型。该模型主要分为特征降维部分，和聚类部分。首先，针对储层预测模型中的聚类算法的参数优化问题，使用混合粒子群对聚类算法的参数进行寻优，基于聚类的目标，设计合理的寻优目标函数，进而提高聚类结果质量。其次，对整个储层预测模型参数优化进行研究，利用混合粒子群对整个模型的所有参数同时寻优。根据特征降维的评价方法，并结合聚类参数优化的目标函数，提出了预测模型参数优化的目标函数，提升模型预测精度。为了解决标签数据不足导致难以构造准确的有监督分类器性数据的问题，本文提出一种自动标签选择方法，构造了基于类别概率和类结构信息的样本聚类标签的可信度评价函数，在此选择可信度高的样本作为分类算法的训练数据基于分类的储层预测模型参数寻优，本文也利用混合粒子群对所有参数同时寻优。

本文的组织结构安排如下：

第一章阐述了本文的研究背景和研究意义，介绍了国内外相关的研究现状，并给出了本文主要的研究内容和结构安排。

第二章介绍了储层预测模型及其参数寻优研究现状。首先，介绍了介绍本研究所采用的三种储层预测模型：FSC（特征选择-聚类组合模型）、FSFEC（特征选择-特征抽取-聚类组合模型），FSFEFSC（特征选择-特征抽取-特征选择-聚类组合模型）。并介绍了储层预测模型中常用的数据挖掘方法。然后分别介绍了储层预测模型中特征降维和聚类分析两部分的参数寻优研究现状及其存

在的问题。最后，对上述存在的问题进行总结归纳，并描述了本文研究技术路线。

第三章介绍基于混合粒子群的聚类算法参数寻优方法。首先，介绍了用于实值空间寻优的基本粒子群算法和用于离散值空间寻优的二进制粒子群算法的基本原理和算法步骤。然后利用混合粒子群对聚类参数进行寻优，并根据聚类目标，设计合适的寻优目标函数。最后，介绍了本研究所用的样本数据、实验环境。对 K 均值和自组织特征映射进行参数优化实验，与没有进行参数寻优的高斯混合模型的聚类结果进行对比，分析实验结果。

第四章介绍了基于混合粒子群的预测模型参数寻优方法。首先，介绍了特征降维几种评价方法。基于特征降维评价方法和聚类目标，设计合理的寻优目标函数，利用混合粒子群算法对整个模型所有参数同时进行寻优。最后对三种常用的储层预测模型进行参数寻优实验，分析实验结果。

第五章介绍了基于类几何结构和类概率的自动标签选取方法。提出了一种聚类结果标签可信度计算方法。在此基础上，按照一定方式选取可信度高的样本及其聚类标签作为训练数据。将基于自动标签的支持向量机和朴素贝叶斯的分类结果与高斯混合模型的聚类结果进行对比，验证自动标签选取方法的有效性。并且也进行了基于分类算法的储层预测模型参数寻优实验，对分类结果进行分析。

第六章总结本文所做的研究内容，并对未来的研究方向进行了展望。

第二章 储层预测模型及其参数寻优研究现状

2.1 储层预测模型介绍

地震数据含有地下岩层构造、岩性和油气储层等地质特征信息^[26]。如今可以从地震数据中提取上百种地震属性，不同的地震属性从不同的角度反映地下地质特征。地震解释其中一个重要的研究方向就是如何利用地震属性进行储层预测。地震属性与地质特征之间通常是非线性关系，并且随着勘探工区和目的地层而变化。并不是所有的地震属性都能反映储层地质特征，甚至有的地震属性会干扰储层预测。可以利用数据挖掘技术中的特征选择方法从原地震属性集合挑选出最佳属性子集，或者使用数据挖掘技术中的特征抽取方法从原地震属性集合中提取最能反映储层地质特征的新属性集合。实际上绝大部分的地震数据都是无标签数据，因此使用数据挖掘技术中的无监督聚类分析方法对地震属性进行分析，在聚类结果的基础上，结合测井资料，找到反映储层地质特征的一类数据，从而建立了地震数据与储层分布之间的关系，正确的反映出地震数据与地下地质特征之间的非线性关系。

图 2-1 列举三种常用的储层预测模型。第一种是特征选择和聚类组合模型（FSC），先利用特征选择方法如方差法，Fisher 分数法从原始地震属性中筛选出重要，有用的属性，抛弃冗余属性，组成最佳地震属性子集。然后聚类算法根据地震性子集含有地质信息将地震数据进行分类，并结合已有的测井数据，分析出类别代表的地质属性。第二种是特征选择，特征抽取和聚类组合模型（FSFEC）。利用特征选择剔除冗余，相关性强的地震属性之后，使用特征抽取在最佳地震属性子集上提取出更能反映地质特征的新地震属性。在新地震属性基础上，使用聚类方法进行地震数据归类。第三种模型是特征选择，特征抽取，特征选择和聚类组合模型（FSFEFSC）。对特征抽取方法提取出的新地震属性再进行筛选，去掉冗余，相关性强的属性，保留有用的属性^[27]。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库