数据挖掘技术及在石油地质中的应用

柴明锐 万成祥 著

天津出版传媒集团
天津科学技术出版社

数据挖掘技术及在石油地质中的应用

柴明锐 万成祥 著

天津出版传媒集团
天津科学技术出版社

图书在版编目（CIP）数据

数据挖掘技术及在石油地质中的应用 / 柴明锐，万成祥著. -- 天津 : 天津科学技术出版社，2017.9
ISBN 978-7-5576-3858-0

Ⅰ. ①数… Ⅱ. ①柴… ②万… Ⅲ. ①石油天然气地质－数据采集 Ⅳ. ①P618.130.2

中国版本图书馆CIP数据核字(2017)第224355号

责任编辑：布亚楠

天津出版传媒集团
天津科学技术出版社出版
出版人：蔡　颢
天津市西康路35号　　邮编 300051
电话：（022）23332695（编辑室）
网址：www.tjkjcbs.com.cn
新华书店经销
北京厚诚则铭印刷科技有限公司印刷
开本 889×1194　1/16　印张 17.0　字数 300 000

2017年9月第1版第1次印刷
定价：45.00元

前 言

近年来，随着油气资源战略的需要以及勘探开发技术的进步，致密油气受到越来越多的关注。

致密油气属于非常规油气的范畴，资源量异常丰富，但其储层测井响应差，横向变化快，平面预测难，一些常规测井评价的方法已经不再适用，因此，急需探索新的技术来应对致密油气带来的挑战。

数据挖掘技术是21世纪信息技术研究的热门问题之一，它作为一个创新性的领域，实用性很强，可以高效地从海量的数据中得到隐藏在数据背后的、有价值的信息，已广泛应用于社会、经济、商业、金融、教育、医学等行业，然而在石油地质领域的应用尚处于初始阶段。

目前，随着石油勘探的不断拓展，石油地质数据以飞快的速度积累，原有的许多应用方法显得无能为力，且低孔低渗等复杂储层评价问题日渐普遍，因此，有必要引入数据挖掘技术。

数据挖掘技术众多，主要包括机器学习、数理统计、数据库、模式识别、粗糙集、模糊数学等相关技术。鉴于机器学习方法实用性最强、应用最为广泛、最易于与石油地质领域结合，本书重点介绍机器学习方法并包括部分统计学方法。

机器学习是在概率论、计算机、数理统计等领域迅速发展起来的交叉学科，可以从数据中学习规律并对未知进行有效的预测，已成为当今的研究热点。

机器学习方法可以通过Matlab、R、Python等来实现，常常涉及复杂的算法与编程，而石油地质工作者编程能力大多较为薄弱，现实情况也不允许花费大量的时间来学习编程，常会望而生畏。“山重水复疑无路，柳暗花明又一村”，SPSS Modeler的出现大大降低了机器学习方法实现的门槛。使用者通过设置节点、调节参数、拖拖拽拽，就能完成专业建模的过程，并且操作界面非常的友好。简言之，该软件在兼顾专业性的同时保持非常友好的易用性，为石油地质工作者在短时间内实现机器学习方法提供了便利。

因此，本书主要介绍如何利用SPSS和SPSS Modeler来实现机器学习方法和部分统计学方法并进行石油地质数据的挖掘。同时，为了使读者尽可能多地了解机器学习方法，作者尽量少地使用数学公式以及计算机编程；为了使读者更高效地掌握机器学习方法，作者重点选

取实用、高性能的算法进行讲解；为了更快地使读者能够利用机器学习方法，作者会详细地介绍对应算法软件的操作。所以，本书非常适合数学、计算机编程涉及较少的石油地质学的本科生、研究生、工作者，以及具有类似情况的对数据挖掘感兴趣的人士。

现在正是技术创新的最佳时机，复杂的地层带给我们很多矛盾现象，这些矛盾就是前人思维中的盲点，也是催生我们技术进步的支点。希望本书能点燃广大读者，尤其是石油地质工作者的火焰，让我们能够更加深入、透彻地认知复杂的地质环境。

此为试读，需要完整PDF请访问：www.ertongbook.com

目 录

1 概述……1

1.1 致密油气带来的挑战与机遇 ……1

1.1.1 致密油气的发展 ……1

1.1.2 致密油气与常规油气的区别及挑战 ……2

1.2 数据挖掘与机器学习 ……4

1.2.1 学习的种类和任务 ……6

1.3 SPSS 和 SPSS Modeler 简介 ……7

1.3.1 SPSS 界面概述 ……8

1.3.2 Modeler 界面概述 …… 11

1.3.2.1 主要节点介绍 …… 13

2 监督学习…… 23

2.1 人工神经网络 …… 23

2.1.1 手把手教你——神经网络 …… 24

2.1.1.1 神经网络界面介绍 …… 24

2.1.2.2 简单例子 …… 32

2.2 支持向量机 …… 42

2.2.1 手把手教你——SVM …… 48

2.2.1.1 SVM 界面介绍…… 48

2.2.1.2 简单例子 …… 50

2.3 决策树 …… 57

2.3.1 C5.0 …… 58

2.3.2 C&RT …… 59

2.3.3 手把手教你——C5.0 …… 59

2.3.3.1 C5.0 界面介绍 …… 59

2.3.3.2 简单例子 …… 63

2.3.4 手把手教你——C&RT …… 70

2.3.4.1 C&RT 界面介绍 …… 70

2.4 贝叶斯网络 …… 85
2.4.1 手把手教你——贝叶斯网络 …… 86
2.4.1.1 贝叶斯网络界面介绍 …… 86
2.4.1.2 简单例子 …… 88
3 非监督学习…… 98
3.1 聚类分析 …… 98
3.1.2 层次聚类分析…… 99
3.1.3 非层次聚类分析……100
3.1.4 聚类分析的优缺点……100
3.1.5 手把手教你——聚类分析 ……100
3.1.5.1 聚类分析界面介绍 ……100
3.1.5.2 简单例子 ……107
3.2 异常值分析 ……110
3.2.1 手把手教你——异常值分析 ……111
3.2.1.1 异常值分析界面介绍……111
3.2.1.2 简单例子 ……114
3.3 因子分析 ……122
3.3.1 因子分析的数学模型 ……122
3.3.2 因子分析的优缺点 ……123
3.3.3 手把手教你——因子分析 ……124
3.3.3.1 因子分析界面介绍 ……124
3.3.3.2 简单例子 ……129
3.4 关联分析 ……133
3.4.1 手把手教你——Apriori 关联分析……135
3.4.1.1 Aprior 关联分析界面介绍 ……135
3.4.1.2 简单例子 ……138
4 集成方法……146
4.1 Boosting ……147
4.2 Bagging ……147
4.3 随机森林 ……148
4.3.1 手把手教你——集成学习 ……148

4.3.1.1 集成算法界面介绍……148
4.3.1.2 简单例子 ……152
5 其他 ……161
5.1 相关性分析 ……161
5.1.1 相关性分析界面介绍 ……162
5.1.1.1 偏相关界面介绍……162
5.1.1.2 双变量相关性界面介绍……165
5.1.2 简单例子……170
5.2 标准化处理 ……173
5.2.1 标准化简介 ……173
5.2.2 手把手教你 ……173
5.3 特征选择 ……175
5.3.1 手把手教你——特征选择 ……177
5.3.1.1 特征选择界面介绍 ……177
5.3.1.2 简单例子 ……179
6 实例……186
6.1 岩性识别 ……186
6.1.1 前言 ……186
6.1.2 原理及实现 ……196
6.1.3 应用实例 ……197
6.1.3.1 标准化处理 ……197
6.1.3.2 异常值处理 ……199
6.1.3.3 优选变量……200
6.1.4 结论 ……206
6.2 沉积体系分析 ……206
6.2.1 前言 ……206
6.2.2 原理及实现 ……207
6.2.3 应用实例 ……207
6.2.3 聚类分析 ……208
6.2.3.1 多期次物源供给分析……208
6.2.3.1.2 沉积体系划分 ……209

6.2.4 物源方向判断 ……213
6.2.5 因子分析——明确母岩类型 ……213
6.2.6 结论 ……216
6.3 储层物性参数预测 ……216
6.3.1 前言 ……216
6.3.2 原理及实现 ……221
6.3.3 应用实例 ……222
6.3.3.1 标准化处理 ……223
6.3.3.2 异常值处理 ……225
6.3.3.3 优选变量 ……226
6.3.3.4 建模及误差分析 ……227
6.3.4 结论 ……231
6.4 储层分类评价 ……231
6.4.1 前言 ……231
6.4.2 原理及实现 ……236
6.4.3 应用实例 ……237
6.4.3.1 标准化处理 ……237
6.4.3.2 异常值处理 ……239
6.4.3.3 建模储层分类模型 ……240
6.4.3.4 分类模型评价 ……242
6.4.3 结论 ……246
6.5 储层影响因素关联分析 ……246
6.5.1 前言 ……246
6.5.2 原理及实现 ……249
6.5.3 应用实例 ……250
6.5.3.1 网络图分析 ……251
6.5.3.2 关联规则建模 ……252
6.5.3.3 关联规则模型分析 ……253
6.5.4 结论 ……255

1 概述

1.1 致密油气带来的挑战与机遇

1.1.1 致密油气的发展

随着世界范围内非常规油气资源勘探热度的不断增加，致密油气已经成为现实的资源接替领域。尤其是致密油，近年来已经成为全球油气勘探的热点领域。贾承造等（2012）提出致密油（图 1-1）是指以吸附或游离状态赋存于生油岩中，或与生油岩互层、紧邻的致密砂岩、致密碳酸盐岩等储集岩中，未经过大规模长距离运移的石油聚集[1]。致密砂岩气是指覆压基质渗透率小于或等 0.1 mD 的砂岩气层，单井一般无自然产能或自然产能低于工业气流下限，但在一定经济条件和技术措施下可获得工业天然气产量。通常情况下，这些措施包括压裂、水平井、多分支井等[2]。

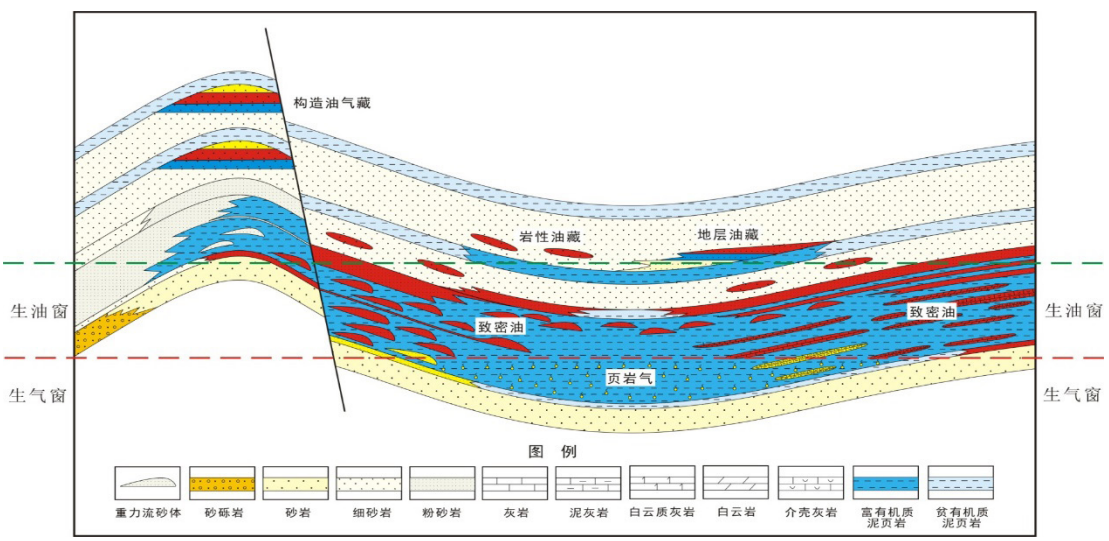

图 1-1 致密油分布模式图

美国是目前世界致密油勘探开发的核心地区。在致密油有效开发的助推下，2008 年美国原油产量止跌回升，扭转了美国持续 24 年以来的石油产量下降趋势。2011 年，美国致密

油产量已达 3 000×10^4 t，2012 年产量高达 9 500×10^4 t，彻底改变了美国能源供应的格局。继巴肯（Bakken）和鹰滩（Eagle Ford）致密油获得重大突破之后，2012 年，美国在加州的蒙特利（Monterey）地区致密油又获战略性突破，引起世界关注。美国能源信息署（EIA，2013）预测，蒙特利地区致密油可采资源量 18.8×10^8 t，约占全美致密油总资源量的 41%，致密油俨然已经成为美国石油能源的有效接替领域。目前我国致密油刚处于起步阶段，近两年来已经在鄂尔多斯、准噶尔、柴达木、松辽、渤海湾、四川、三塘湖等盆地取得了实质性进展，发现了鄂尔多斯长 7 致密油、准噶尔昌吉大油田、松辽盆地扶杨油层等重点领域，初步落实储量规模达 23.3×10^8 t，有望成为我国“十三五”能源储备的主要战略领域。

关于致密砂岩气的研究，始于 20 世纪 50 年代。从圣胡安盆地的隐蔽气藏，到加拿大阿尔伯达盆地埃尔姆沃斯（Elmworth）巨型深盆气藏，再到 Raton 盆地的盆地中心气、致密砂岩气与连续型天然气聚集，目前致密气已成为全球非常规天然气勘探的重点领域。北美已实现致密砂岩气的大规模商业化生产，2010 年美国致密气产量达 1 754×10^8 m^3，占当年天然气总产量 6 110×10^8 m^3 的 30%。中国致密砂岩气的发展非常快，2011 年苏里格致密砂岩大气区实现探明储量超 3.0×10^{12} m^3，约 3500 口生产井，单井平均产量 1.0×10^4 m^3/d，年产量 135×10^8 m^3；四川盆地须家河组致密砂岩大气区资源量超过 5.0×10^{12} m^3，已发现三级储量 1.0×10^{12} m^3，产量超过 30×10^8 m^3。此外，松辽盆地登娄库组、吐哈盆地水西沟群、准噶尔盆地八道湾组、塔里木盆地库车东部侏罗系及西部深层巴什基奇克组致密砂岩均具备致密气区的条件，取得重大发现。2011 年，致密气储量已占中国天然气的三分之一以上，年产量超过 200×10^8 m^3，展现出很大的潜力，可采资源量（15 ～ 20）×10^{12} m^3。

1.1.2 致密油气与常规油气的区别及挑战

随着石油工业的不断发展，油气勘探与开发从常规油气领域跨向非常规油气领域，这也是一种必然的趋势。但是两者在地质特征、聚集特征及成藏机理等方面有着很大的区别。常规油气研究的灵魂是成藏，目标是回答圈闭是否有油气；非常规油气研究的灵魂是储层，目标是回答储集有多少油气[2]。非常规油气主要包括致密油气、页岩油气、煤成气、重油沥青及天然气水合物等，其中，致密油气与页岩油气是当今非常规油气勘探开发的重要领域。我国目前非常规油气的勘探开发整体上仍处于起步阶段，页岩气取得了一定的突破，处于示范区突破和建产阶段，但致密油气还停留规模勘探开发的初期。

实际勘探过程中发现，致密油气与常规油气在分布聚集上首先就有很大的差别（表 1-1），常规油气藏主要发育在断陷盆地大型构造带、前陆冲断带大型构造、被动大陆边缘以及克拉通大型隆起等正向构造单元中，具有常规二级构造单元控制油气分布的特征。油气聚集于构造高点或岩性圈闭和地层圈闭中，平面上呈孤立的单体式分布或较大规模的集群式分布。

此为试读，需要完整PDF请访问：www.ertongbook.com

致密油气的分布更加复杂，油气并不局限于二级构造单元，主要分布于前陆盆地坳陷—斜坡、坳陷盆地中心及克拉通向斜部位等负向构造单元，涵盖了盆地中心及斜坡，有效勘探范围扩展至全盆地，油气大面积分布，丰度不均一。平面上，多层系叠合连片含油；纵向上，油气滞留在烃源岩内或连续分布于紧邻烃源岩上下的大面积致密储层中，从而形成大规模展布的油气聚集。致密油气流体分异差，无统一的油水界面，油、气、水常多相共存，同时含油气饱和度变化大，具有“满凹含油气”的特征。

表 1-1 致密油气与常规油气聚集特征对比

特征	常规油气聚集（非连续圈闭聚集）	致密油气聚集（连续型储层聚集）
聚集单元	构造、岩性与地层等常规圈闭	无明显界限的非闭合圈闭
储层特征	毫米－微米级孔喉	纳米级孔喉，滞留作用明显
源储配置	一般源外成藏，排聚时刻匹配	大面积源储紧邻及共生
水动力作用	明显，重力分异、浮力聚集	不明显，流体分异差、浮力作用受限
运移方式	远距离二次运移为主	一次运移或短距离运移
渗流机理	达西渗流或管流	滞留、非达西渗流为主
油气水关系	上油气下水、压力系统界面明显，单井有自然产量	无统一油气水界面与压力系统，饱和度差异大，一般油气水共存、单井无自然工业产量
分布和聚集	单体型、集群型非连续分布，局部富集	盆地中心、斜坡等大面积（准）连续型分布，有“甜点”区
技术应用	直井、酸化压裂等常规技术	水平井支井、分段分层压裂等特殊技术

此外，致密油气储层特征也是致密油气研究过程中的一项重要内容。致密油气储层既可以是致密砂岩，也可以是致密碳酸盐岩，但是相对于常规储层来说，致密油储层的物性条件要相差甚远，而且在储层中寻找油气“甜点”的思路也大不相同。通过分析国内外典型盆地致密油的储层物性发现（表 1-2），一般致密油储层孔隙度多小于 12%，覆压渗透率小于 0.1×10^{-3} μm^2，空气渗透率小于 1×10^{-3} μm^2。与常规储层不同，致密储层为覆压基质渗透率孔隙度≤ 10% 的砂岩、灰岩储层，孔喉以微孔、纳米孔为主，微孔直径均 ＜2 μm，主体直径为 40 ～ 900 nm，主要为微米－纳米级孔喉（孔喉直径在 2.0 ～ 0.03 μm 之间）网络系统。中国致密储层中 77.2% 的中值孔喉半径 ＜ 1 μm，属于纳米级孔喉，特别是在鄂尔多斯盆地苏里格地区致密储层气储集层孔喉半径＜ 1 μm 的占到了 83.6%[3]。结合当前我国致密油勘探开发过程中的有效储集空间，在目前勘探方法和技术条件下，致密油的“甜点”区有效孔隙度大于 5% ，渗透率在不同探区的评价范围不一致 [4]。

表 1-2 国内外主要盆地致密油储层物性数据对比

地区	盆地名称	层位	孔隙度 /%	空气渗透率 /（10^{-3} μm^2）	“甜点”区孔隙度 /%	“甜点”区空气渗透率 /（10^{-3} μm^2）
中国	鄂尔多斯	三叠系延长组 6—7 段	2.0 ～ 12.2	0.01 ～ 1.14	5 ～ 12	0.30 ～ 1.14
	四川	侏罗系砂岩	＜ 6	0.001 ～ 1		
	松辽	青山口组、扶杨油层	2 ～ 15	0.6 ～ 1.0	6 ～ 15	0.6 ～ 1
	渤海湾	沙河街组	5 ～ 15	0.3 ～ 4.6	6 ～ 15	0.3 ～ 4.6
	准噶尔	二叠系芦草沟组	3 ～ 16	＜ 1	5 ～ 16	0.5 ～ 1
	柴达木	第三系下干柴沟组砂岩	3 ～ 8	0.05 ～ 1	5 ～ 8	
	三塘湖	二叠系芦草沟组、条湖组	2 ～ 12	＜ 0.005 ～ 2	5 ～ 12	0.5 ～ 1
北美	威利斯顿	上泥盆统巴肯组	2 ～ 13	＜ 0.01 ～ 0.1	5 ～ 13	0.3 ～ 1
	南德克萨斯	白垩系鹰滩组	2 ～ 12	＜ 0.01 ～ 0.1	5 ～ 12	0.3 ～ 1

综上所述，致密油气分布更加复杂，油气的分布并不局限于二级构造带，平面上，多层系叠合连片含油；纵向上，油气滞留在烃源岩内或连续分布于紧邻烃源岩上下的大面积致密储层中。致密储层的物性普遍很差，在大面积连续分布的致密储层中寻找“甜点”，更是难上加难。因此，对致密油气这类数据量更大但规律并不明显的非常规资源来说，大数据分析和计算机学习的方法会对致密油气的勘探与开发提供巨大的帮助！

1.2 数据挖掘与机器学习

随着数据大爆炸时代的到来，人类在各个领域积累了大量的数据，可是在数据分析处理方面缺乏有效的手段，人们很难在海量数据下挖掘隐藏的有价值的信息，信息爆炸演变成了“信息坟墓”（图 1-2）。早在 1982，John Naisbitt 就已预言：“人们将处于信息丰富却知识贫乏的孤岛上。”怎样将这些庞大的数据转变成有用的信息和知识成了当代人们迫切的需要。在这种背景下，数据挖掘充分吸取了机器学习、高性能计算、信息提取、数理统计、数据可视化、模式识别、数据库、人工智能等多个领域的养分，于 20 世纪 80 年代迅速发展起来，并快速渗透到商业、金融等各个行业中，作为一个新兴的领域登上历史的舞台。

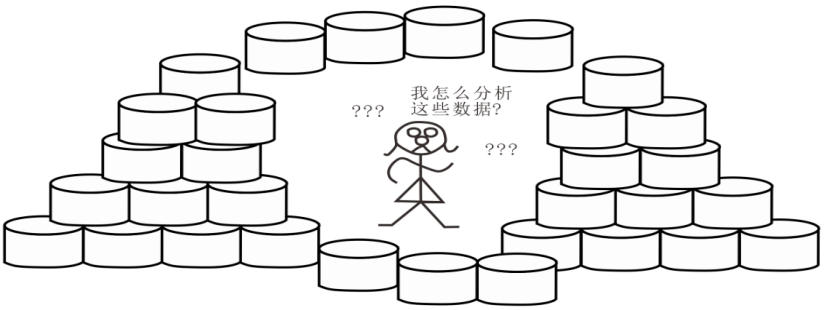

图 1-2 我们处于数据丰富但无法获取有用信息的窘境

数据挖掘指的是从大量的、残缺的、有干扰的数据中提取隐含的、有效的、有价值的且最终可理解的模式的过程[5]（图 1-3）。其具体的挖掘过程包括以下 6 个阶段：①业务理解；②数据预处理；③特征提取；④特征选择；⑤建模和模型评估；⑥模型应用（图 1-4）。

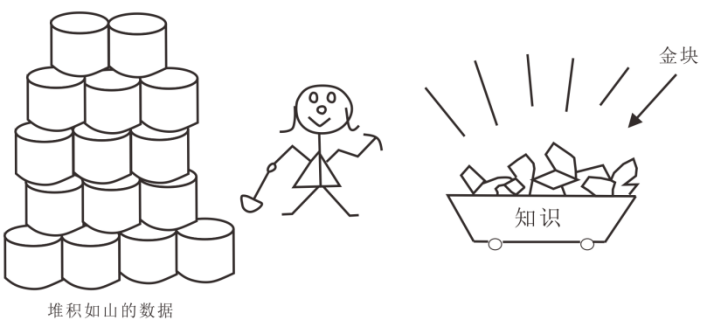

图 1-3 数据挖掘示意图示（利用现有的数据挖掘出有用的知识）

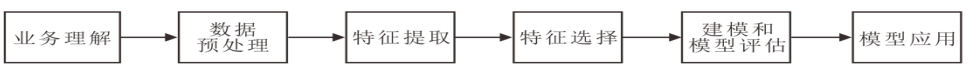

图 1-4 数据挖掘流程图

数据挖掘依靠的技术繁多，其核心主要是依靠机器学习的算法来实现。

机器学习始于20世纪50年代，是一门“古老”而又“新兴”的计算机科学技术，属于人工智能（Artificial Intelligence）研究的分支，具体指的是设计一种让计算机能够自动学习的算法，从而使计算机从大量数据中获得隐藏的规律，并利用获得的规律对未知的情况进行预测。

机器学习具有以下特点：①大多数机器学习所解决的问题都不能使用固定的流程代码实现，而这些问题对人类而言却十分简单；②学习能力的程序指的是可以不断地从数据和经历中吸取经验教训，对未来进行预测；③机器学习具有能够不断改善自身从而应对具体任务的能力。

机器学习技术实用性强，在各个行业得到了广泛的应用，但是在石油地质领域的应用尚处于探索的阶段，主要应用于岩性、孔隙流体以及储层物性等方面的识别[6-10]。机器学习方法在石油地质领域的应用中，很大程度上是成功的，但是也存在很多不成熟的地方：主要体现在实际应用中，过多强调了优化算法的重要性，却忽略了输入变量的选择问题，无论算法多么优越，与预测目标相关性不强变量的引入，势必会对预测精度产生巨大的影响；过多强调了单一算法的优越性，但是由于数据特征千差万别，即使再好的算法也无法适用于所有的预测问题中；应用范围不够广泛。

其原因在于，机器学习算法理论性强，但实践性弱。《孙子兵法·作战》有云，“不尽知用兵之害者，则不能尽知用兵之利”，意思是说，不对用兵打仗的坏处与弊端进行充分的了解，不可能对用兵打仗的好处有足够的认识。算法的应用是一个辩证的过程，不仅在于不同算法间的比较和搭配使用有着辩证的关系，在同一个算法中，不同参数和阀值的设置同样会带来大相径庭的结果。

Aminzadel（2005）指出，数据挖掘技术是将石油勘探数据转化为有用的信息和知识的关键[11]。致密储层复杂性强，如何有效地将机器学习方法应用其中，将成为打开致密储层神秘面纱的一个关键因素。

1.2.1 学习的种类和任务

机器学习技术主要包括监督学习（supervised learning）、无监督学习（unsupervised learning）、集成学习（Ensemble Learning）。其中，监督学习关注的是对未知情况的预测，主要包括分类问题（classification）和回归问题（regression）；无监督学习更关注的是分析事物本身的特性（图1-5），使用的技术有降维（dimensionality reduction）、聚类（clustering）、关联分析（correlation analysis）等；集成学习则关注的是对未来预测的准确性，也就是提高机器学习的泛化能力，使用的技术包括Boosting、Bagging、随机森林等。

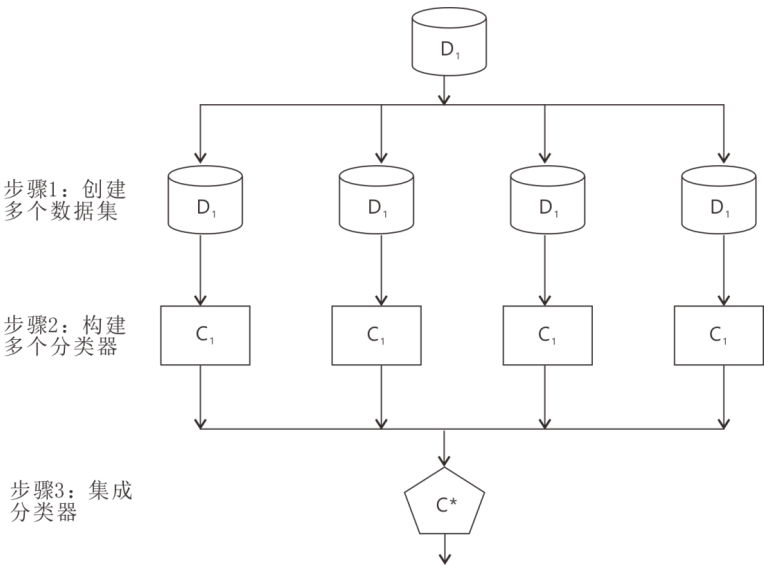

图 1-5 集成学习逻辑视图

分类问题，属于预测问题，其预测目标变量的类型属于分类型。比如，预测致密砂砾岩储层的岩性就是一个分类问题。

回归问题也是属于预测问题，其预测目标变量的类型属于连续类型。比如，预测致密砂砾岩储层的孔隙度就是一个回归问题。

聚类分析指的是依据样本间的亲属性，进行分类，把亲属性强的聚为一类。需要说明的是，在聚类分析的过程中，我们无法事先就知道了聚类的数量以及每一类所代表的含义。比如，重矿物的聚类分析可以进行沉积体系的划分，亲属性强的重矿物样本被认为属于同一个沉积体系。

降维分析指的是把高维度变量数据通过一定的方法，转化为低维度变量数据，从而提高模型的运算速度和效率。

关联分析指的是挖掘事物之间的相互关系。这种方法在市场营销中应用比较普遍，典型的例子就是“啤酒和尿布”。

1.3 SPSS 和 SPSS Modeler 简介

SPSS（Statistical Product and Service Solutions），即“统计产品与服务解决方案”，由美国斯坦福大学的三位研究生于 20 世纪 60 年代末研制，是世界上最早、应用最广泛的专业统计和数据模型软件之一，主要针对经济、管理、医学、农学、教育、市场研究、社会调查等多个行业和领域。

2009 年，IBM 收购了 SPSS 数据分析软件公司，并将其广受赞誉的 SPSS 统计分析软

件和 Clementine 数据挖掘软件进行整合，成为 IBM 软件部的一个产品，原先的 SPSS 统计分析软件已经成为 IBM SPSS Statics（简称 SPSS），它定位于标准的统计分析需求，而 Clementine 所具有的数据分析与挖掘的功能则由 IBM SPSS Modeler（简称 Modeler）来实现。

具体来说，SPSS Statistics 是一款统计分析软件，提供了执行全程分析所需的核心功能；它易于使用，包含范围广泛的过程和技术，能帮助您增加收入，超越竞争对手，组织调研并制定更好的决策。

IBM SPSS Modeler 一个预测性分析平台，能够为个人、团队、系统和企业做决策提供预测性智能；它不但集成了诸多计算机科学中机器学习的优秀算法，同时也综合了一些行之有效的统计分析方法，成为内容最全面、功能最强大、使用最方便的数据挖掘软件产品。

IBM SPSS Statics 和 IBM SPSS Modeler 保持了 SPSS 产品的一贯作风，功能强大、界面友好、操作简单，具有较高的使用性，受到越来越多的数据分析工作者集科研人员的青睐，但是在石油地质领域中的应用不是那么广泛，尤其是 IBM SPSS Modeler 应用得较少。鉴于石油地质工作普遍缺少较强的数学、计算机编程功底，因此，本书引入 IBM SPSS Statics 和 IBM SPSS Modeler 软件，力求使石油地质工作者便捷地实现数据挖掘技术。

1.3.1 SPSS 界面概述

本次主要介绍的是 IBM SPSS Statics 20.0。

启动 IBM Statics 20.0

要启动此应用程序，请单击：

开始 > 所有程序 >IBM SPSS Statics 20.0>IBM SPSS Statics 20.0。

显示初始化界面，主窗口将在几秒钟后显示（图 1-6）。

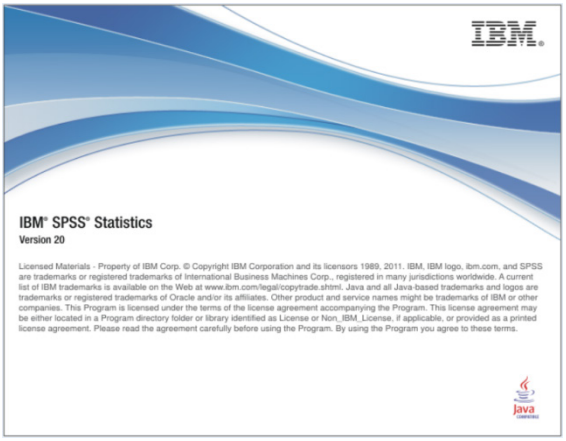

图 1-6 IBM SPSS Statics 20.0 初始化界面

此为试读，需要完整PDF请访问：www.ertongbook.com