

Lecture 2

模式识别系统举例

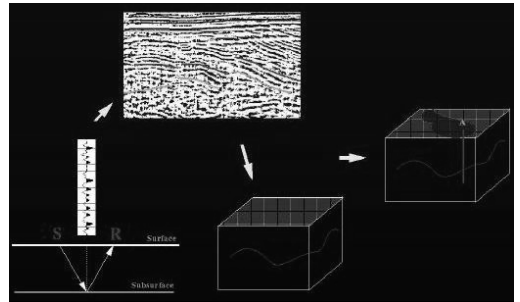
continued

Example Application of PR in Geophysical Exploration

简介：石油的形成与开采

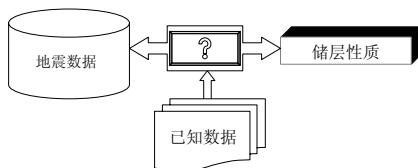


石油地球物理勘探原理简介

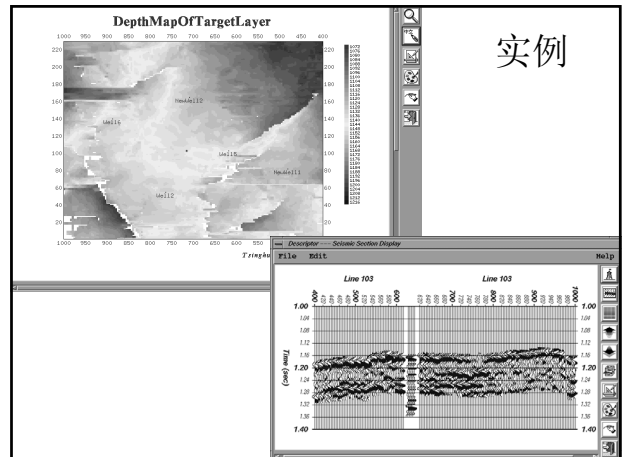


根据地震数据进行储层分析 ——石油物探中的模式识别问题

- 地震反射的到达时间
——反映地层的构造信息
- 地震数据的波形、频率成分
——与地层的岩性、含油气性质有关



实例



实例（续）

• 方案1. 监督模式识别

- 提取与储层性质关系密切的地震数据特征
 - 振幅、能量、频率、相位…
- 将待研究的储层性质归纳为几个类别
- 用已知井（旁的地震数据）作为训练样本
- 得到某种分类器，对未知位置进行分类预测

• 问题：

- 训练样本过少
- 训练样本缺乏充分的代表性

实例（续）

• 方案2. 非监督模式识别

- 提取与储层性质关系密切的地震数据特征
- 对这些数据进行聚类分析
- 考察聚类结果，根据有关地质知识和已知井信息判断结果的合理性，对聚类进行解释
- 根据聚类结果对未知位置进行判断

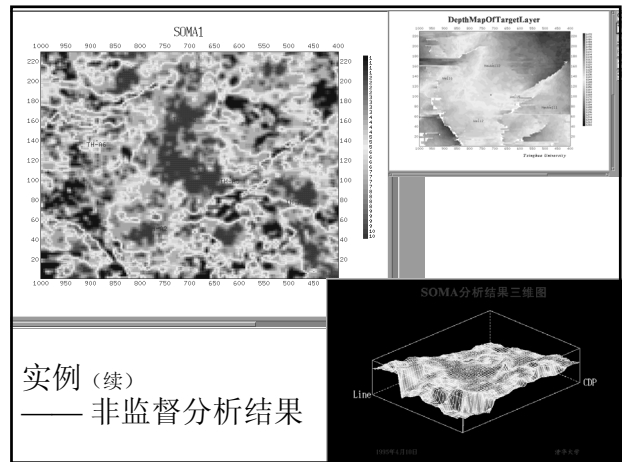
• 问题：

- 只能得到定性结果，对已知井的信息利用不充分

实例（续）

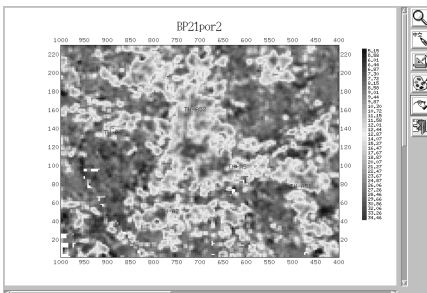
• 方案3. 非监督模式识别+监督模式识别

- 首先进行非监督模式识别分析
- 根据非监督分析结果指定若干虚拟训练样本
- 用已知井和虚拟训练样本组成训练样本集
- 利用监督模式识别再次进行学习和预测



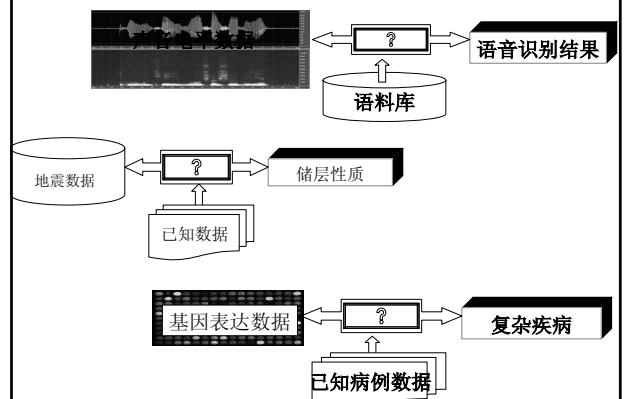
实例（续）

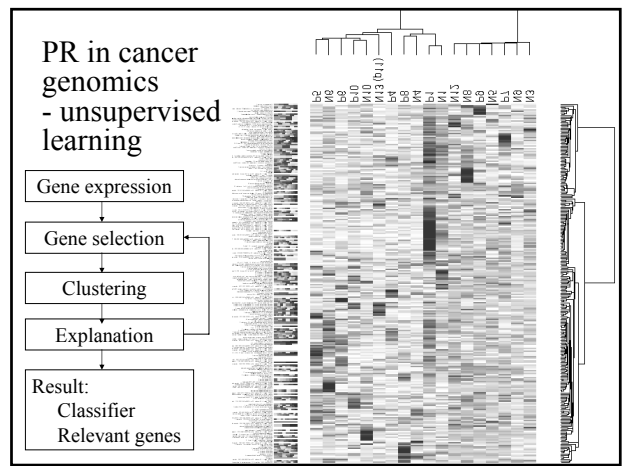
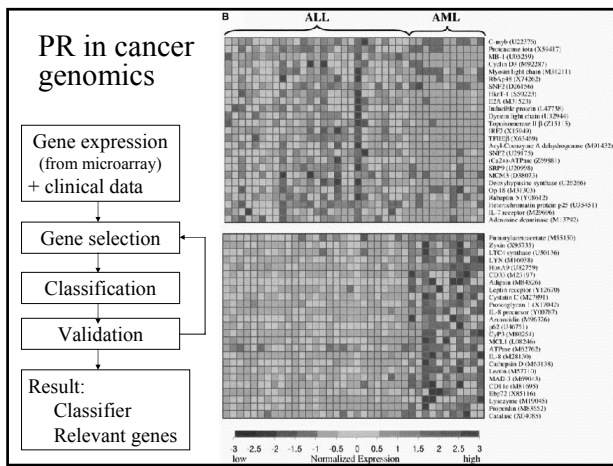
——非监督分析基础上，定量监督学习后的预测结果



Wells	Depth(m)	Sand Thickness	
		Actual	predicted
TH-A2	1233-1410	35.7m	known
TH-A5	1242-1420	51.9m	known
TH-A6	1239-1416	21.8m	known
TH-A51	1264-1441	42.1m	42.4m
TH-A52	1261-1440	45.6m	43.8m

模式识别属于基于数据的机器学习范畴





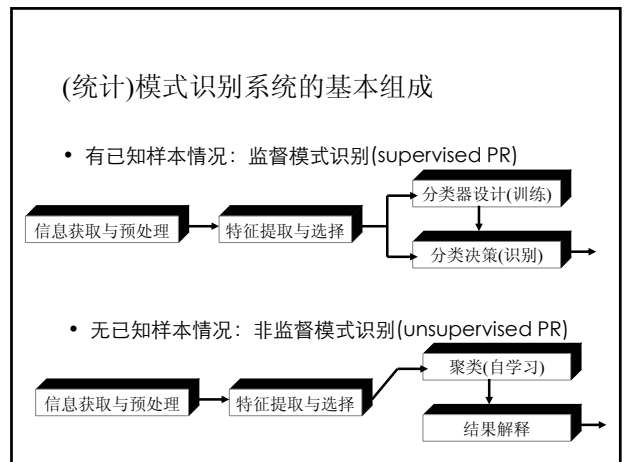
从这些例子我们看到了什么？

特征

- 特征很重要
 - 领域知识
 - 特征选择
 - 特征提取

分类和聚类

- 分类和聚类是模式识别系统的核心
 - 依据概率分布模型
 - 直接依据数据
- 监督学习
 - 分类器
 - 对样本的要求/样本的选择
- 非监督学习
 - 聚类方法
 - 结果的解释



处理监督模式识别问题的一般步骤

1. **分析问题**: 看是否属于模式识别问题, 把研究的目标抽象为类别; 分析问题中哪些(可以观测的)因素可能与分类有关
2. **原始特征获取**: 设计实验方法, 得到已知样本, 对这些样本实施观测和预处理, 获取与样本分类有关的观测向量(原始特征)
3. **特征选择与提取**: 为了更好地进行分类, 对特征进行必要的提取与选择
4. **分类器设计**: 利用已知样本设计(训练)某种分类器
5. **分类**: 对未知样本, 实施同样对信息获取与预处理、特征提取与选择, 用设计对分类器进行识别

处理非监督模式识别问题的一般步骤

1. **分析问题**: 看研究对目标是否可能抽象为若干类别; 分析问题中哪些(可以观测的)因素可能与所关心对类别有关
2. **获取原始观测**: 观测未知样本, 获取原始特征
3. **特征提取与选择**: 进行必要的特征提取与选择
4. **聚类分析**: 采用某种方法将未知样本分类
5. **结果解释**: 分析所得的类别与所关心的目标之间对关系; 如问题需要, 用同样的方法对新的未知样本进行分类