|  |  |  |  |
| --- | --- | --- | --- |
| 第一章概念：**模式**：指可测量且需要分类的对象的描述。**模式类**：把模式所属的类别或同一类别中模式的总体 称为模式类（或简称为类）。**样本**：所研究对象的一个个体。**样本集**：若干样本的集合。**特征**：指用于表征样本的观测，通常是数值表示的某 些量化特征。**➢** 如果存在多个特征，则它们就组成了特征向量。➢ 样本的特征组成了样本的特征空间，空间的维数就是特征的个数，而每个样本是 特征空间中的一个点。➢ 某些情况下，对样本的原始描述可能是非数值形式的，需要采用一定的方法把这 些特征转换成数值形式。**模式的基本属性：{可观测性**：待处理模式可由某种类型的传感器装置获取， 这样的传感装置可以是基于某种物理效应的传感器，也可以是基于某种化学效应或生物效应的传感器，甚至可以是 某种基于数学模型的虚拟传感器。**可区分性**：不同模式类的观测样本之间应该具有可区分的特征。**相似性**：同一模式类的观测样本之间应该具有某种相似的特征。}**识别**：再认知，把对象分门别类的认出来。**模式识别：**用计算的方法根据样本的特征将样本划分到一定的类别中去。侧重于对人的认知行为进行模仿，把人的知识和经验转化为可以为机器所利用的一些规则和方法，赋予机器对被观测事物进行综合分析和自动分类的能力，使机器可以根据被观测事物过往的观测样本形成相应的分类规则并据此完成对新的观测样本进行分类的任务。**难点**：样本中不仅包含待识别事物的固有信息，也包含有害的环境信息。**模式采集**：模式的数字化表达。**若干问题：**1)紧致性 (2)特征选取 (3)相似性度量与分类 (4)性能评价。 |  |  |  |
|  |  |  | **分级聚类方法**：思想：从各类只有一个样本点开始，逐级合并，每级只合并两类，直到最后所有样本都归到一类。  **算法步骤（从底向上）：**（1）初始化，每个样本形成一类（2）把相似性最大（距离最小）的两类合并（3）重复（2），直到所有样本合并为两类。常见的几种类间似性度量：最近距离，最远距离，均值距离  模糊特征的分类：  特征的模糊化：基本思想：根据一定的模糊化规则把原来的一个（或几个）特征变量分成多个模糊变量，使每个模糊特征表达某一局部特性，利用这些新特征来进行模式识别。  结果的模糊化：概念：使分类器输出不是硬分类，而是给出属于各类的程度。  优点：1.更好表达分类结果中的不确定性因素，利于根据结果进行决策。2.利于后期进一步处理和分析。  结果模糊化的方法：结合知识和所用分类器确定。如：1.依据样本离类别中心的距离，离分类面的距离，2.与已知样本或类别中心之间相似性度量，3.神经网络输出值的大小  模糊C均值聚类算法（FCM）：  缺点：        改进的模糊C 均值算法：  特点：AFC有更好的鲁棒，且对给定的聚类数目不十分敏感。但有时可能会出现一个类中只包含一个样本的情况，可通过在距离计算中引入非线性，使之不会小于某个值来改进。AFC、FCM与C均值一样，依赖于初值。改进的模糊C均值算法较前面提到的模糊C均值算法具有更好的鲁棒性，它不但可以在有孤立样本存在的情况下得到较好的聚类效果，而且可以放松隶属度条件，而且因为放松了隶属度条件，使最终聚类结果对预先确定的聚类数目不十分敏感。与确定性C均值算法和模糊C均值算法一样，改进的模糊C均值算法仍然对聚类中心的初值十分敏感，为了得到较好的结果，可以用确定性C均值算法或模糊C均值算法的结果作为初值。  模糊k近邻分类器：k近邻的一个问题：当样本较稀疏时，只考虑样本近邻顺序而不考虑距离远近是不适当的。          **十一、Recurrent Neutral Networks**  处理序列  也叫做序列到序列问题，用编码器-解码器模型解决  基本思想：RNN中含有一个随着序列的产生而更新的内部状态  RNN的优点：1.可以处理任何长度的输入；2.步骤t的计算可以（理论上）使用许多步骤后的信息；3.更长的输入不会增加模型大小；4.每个时间步骤都应用相同的权重，因此处理输入的方式是对称的。  RNN的缺点：1.递归计算很慢；2.在实践中，很难从许多步骤后访问信息  LSTM架构：使RNN更容易在许多时间步长上保存信息，例如，如果f＝1和i＝0，则该cell的信息被无限期地保存。相比之下，普通RNN更难学习在隐藏状态下保存信息的递归权重矩阵Wh。隐藏状态ℎ𝑡 : 短期记忆；𝑐𝑡 : 长短期记忆网络参数𝑊: 长期内存。LSTM不能保证没有消失/爆炸的梯度，但它确实为模型学习长距离依赖关系提供了一种更容易的方法  总结：RNN在架构设计中具有很大的灵活性。原始RNN很简单，但效果不太好。LSTM或GRU很常见：它们的加性相互作用改善了梯度流。RNN中梯度的反向流可能会爆炸或消失。分解由渐变剪裁控制。消失是由加性相互作用（LSTM）控制的。更好/更简单的架构是当前研究的热点，也是对序列进行推理的新范式。需要更好的理解（理论上和实验上）。  **十二、Attention and Transformers**  RNN与Transformer比较：  RNN：  （+）LSTM可以很好地用于长序列。  （-）期望输入的有序序列  （-）顺序计算：只有在完成前一个隐藏状态之后才能计算后续隐藏状态。  Transformers  （+）擅长长序列。每个注意力计算都会考虑所有输入。  （+）可以对无序集或具有位置编码的有序序列进行操作。  （+）并行计算：所有输入的所有对齐和注意力得分都可以并行完成。  （-）需要大量内存：需要为单个自注意力头计算和存储N x M个对齐和注意力定标器。（但GPU越来越大、越来越好）  总结：向RNN添加注意力使其能够在每个时间步长“关注”输入的不同部分；一般注意力层是一种新型的层，可用于设计新的神经网络架构；Transformers是一种使用自我注意力和层规范的层。 |