

2.2 模式识别的算法体系



模式识别技术从 1929 年发端，至今已有快 100 年的历史了。在漫长的发展过程中，随着计算机技术、神经科学、统计科学、语言学等学科的发展，以及工程实践中不断地研究总结，形成了分支庞杂、丰富多样的算法体系。这些算法虽然出发点有所不同，理论基础和算法思路也各有特色，但都可归纳于统计模式识别和结构模式识别两大类。

统计模式识别是主流的模式识别方法，它是将样本转换成多维特征空间中的点，再根据样本的特征取值情况和样本集的特征值分布情况确定分类决策规则。线性分类器是最基本的统计分类器，它通过寻找到线性分类决策边界来实现特征空间中的类别划分。贝叶斯分类器也是统计分类器，它的分类决策规则是基于不同类样本再特征空间中的概率分布，以逆概率推理的贝叶斯公式来得到类别划分的决策结果。最近邻分类器的把学习过程隐藏到了分类决策过程中，通过寻找训练集中与待分类样本最相似的子集来实现分类决策。神经网络分类器来源于对生物神经网络系统的模拟，它的本质是高度非线性的统计分类器，并且随着计算机技术的发展从浅层网络向深度学习不断演化，目前已称为新一轮人工智能热潮的基础。聚类分析是无监督学习的典型代表，目前多采用统计学习方法。

模糊模式识别不是一套独立的方法，而是将模糊数学引入模式识别技术后，对现有各种算法的模糊化改造，它在更精确地描述问题和更有效地得出模式识别结果方面都有许多有价值的思路。

特征降维也不是独立的模式识别算法，但是是完成模式识别任务的流程中不可缺少的的一个步骤。特征降维通过寻找数量更少，对分类更有效的特征来提升整个模式识别系统的性能。

结构模式识别与统计模式识别有根本性的不同，它抽取的不是一系列数值型的特征，而是将样本结构上的某些特点作为类别和共同的特征，通过结构上的相似性来完成分类任务。句法模式识别利用了形式语言理论中的语法规则，将样本的结构特征转化为句法类型的判定，从而实现模式识别的功能。聚类分析中，也可以采用结构特征上的相似性来完成样本类别的划分。