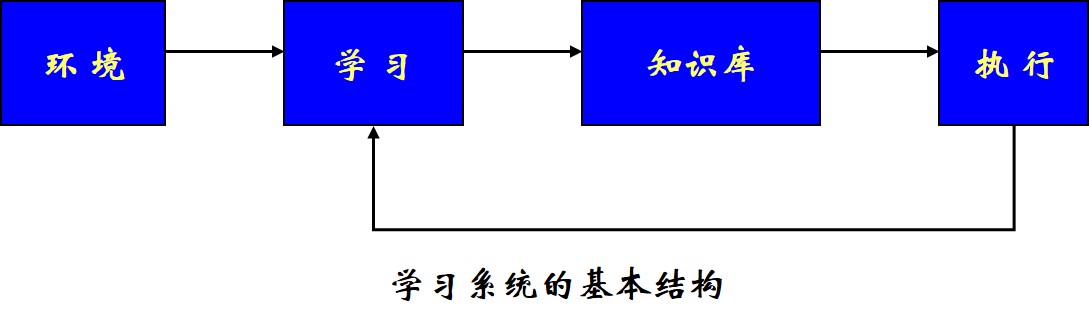
第七章 机器学习

**7-1  什么是学习和机器学习?为什么要研究机器学习?**

答：按照人工智能大师西蒙的观点，学习就是系统在不断重复的工作中对本身能力的增强或者改进，使得系统在下一次执行同样任务或类似任务时，会比现在做得更好或效率更高。 机器学习是研究如何使用机器来模拟人类学习活动的一门学科，是机器学习是一门研究机器获取新知识和新技能，并识别现有知识的学问。这里所说的“机器”，指的就是计算机。 现有的计算机系统和人工智能系统没有什么学习能力，至多也只有非常有限的学习能力，因而不能满足科技和生产提出的新要求。

**7-2  试述机器学习系统的基本结构，并说明各部分的作用。**

答：



环境向系统的学习部分提供某些信息，学习部分利用这些信息修改知识库，以增进系统执行部分完成任务的效能，执行部分根据知识库完成任务，同时把获得的信息反馈给学习部分。 影响学习系统设计的最重要的因素是环境向系统提供的信息。更具体地说是信息的质量。

**7-3  试解释机械学习的模式。机械学习有哪些重要问题需要加以研究?**

答：机械学习是最简单的机器学习方法。机械学习就是记忆，即把新的知识存储起来，供需要时检索调用，而不需要计算和推理。是最基本的学习过程。任何学习系统都必须记住它们获取的知识。在机械学习系统中，知识的获取是以较为稳定和直接的方式进行的，不需要系统进行过多的加工。 要研究的问题：

(1) 存储组织信息

只有当检索一个项目的时间比重新计算一个项目的时间短时，机械学习才有意义，检索的越快，其意义也就越大。因此，采用适当的存储方式，使检索速度尽可能地快，是机械学习中的重要问题。

(2) 环境的稳定性与存储信息的适用性问题

机械学习基础的一个重要假定是在某一时刻存储的信息必须适用于后来的情况

(3) 存储与计算之间的权衡

如果检索一个数据比重新计算一个数据所花的时间还要多，那么机械学习就失去了意义。

**7-4  试说明归纳学习的模式和学习方法。**

答：归纳是一种从个别到一般，从部分到整体的推理行为。 归纳学习的一般模式为：

给定：观察陈述(事实)F，假定的初始归纳断言(可能为空)，及背景知识 求：归纳断言(假设)H，能重言蕴涵或弱蕴涵观察陈述，并满足背景知识。

学习方法

(1) 示例学习

它属于有师学习，是通过从环境中取得若干与某概念有关的例子，经归纳得出一般性概念的一种学习方法。示例学习就是要从这些特殊知识中归纳出适用于更大范围的一般性知识，它将覆盖所有的正例并排除所有反例。

 (2) 观察发现学习

它属于无师学习，其目标是确定一个定律或理论的一般性描述，刻画观察集，指定某类对象的性质。它分为观察学习与机器发现两种，前者用于对事例进行聚类，形成概念描述，后者用于发现规律，产生定律或规则。

**7-5  什么是类比学习？其推理和学习过程为何？**

答：类比是一种很有用和很有效的推理方法，它能清晰，简洁地描述对象间的相似性，是人类认识世界的一种重要方法。

类比推理的目的是从源域S中，选出与目标域T最近似的问题及其求解方法，解决当前问题，或者建立起目标域中已有命题间的联系，形成新知识。

类比学习就是通过类比，即通过对相似事物加以比较所进行的一种学习。 类比推理过程如下：

 (1) 回忆与联想

通过回忆与联想在源域S中找出与目标域T相似的情况。

(2) 选择

从找出的相似情况中，选出与目标域T最相似的情况及其有关知识。

(3) 建立对应关系

在源域S与目标域T之间建立相似元素的对应关系，并建立起相应的映射。

(4) 转换

把S中的有关知识引到T中来，从而建立起求解当前问题的方法或者学习到关于T的新知识。

类比学习过程主要包括：

1. 输入一组已经条件(已解决问题)和一组未完全确定的条件(新问题)
2. 按照某种相似性的定义，寻找两者可类比的对应关系

(3) 根据相似变换的方法，建立从已解决问题到新问题的映射，以获得待求解问题所需的新知识。

(4) 对通过类比推理得到的关于新问题的知识进行校验。验证正确的知识存入知识库中，暂时无法验证的知识作为参考性知识，置于数据库中。

**7-8  用C语言编写一套计算机程序，用于执行BP学习算法。**

**7-9  试应用神经网络模型优化求解销售员旅行问题。**

**7-10 考虑一个具有阶梯型阈值函数的神经网络，假设**

**(1) 用一常数乘所有的权值和阈值；**

**(2) 用一常数加于所有权值和阈值。 试说明网络性能是否会变化?**

答：(1) 不会 (2) 会

**7-11 增大权值是否能够使BP学习变慢?**

答：是。

**7-12 什么是知识发现？知识发现与数据挖掘有何关系？**

答：根据费亚德的定义，数据库中的知识发现是从大量数据中辨识出有效的，新颖的，潜在有用的，并可被理解的模式的高级处理过程。

数据挖掘是知识发现中的一个步骤，它主要是利用某些特定的知识发现算法，在一定的运算效率内，从数据中发现出有关的知识。

**7-13 试说明知识发现的处理过程。**

答：费亚德的知识发现过程包括

(1) 数据选择

根据用户需求从数据库中提取与知识发现相关的数据

(2) 数据预处理

检查数据的完整性与数据的一致性，对噪音数据进行处理，对丢失的数据利用统计方法进行填补，进行发掘数据库

 (3) 数据变换

利用聚类分析和判别分析，从发掘数据库里选择数据

 (4) 数据挖掘

(5) 知识评价

对所获得的规则进行价值评定，以决定所得到的规则是否存入基础知识库

知识发现的全过程，可进一步归纳为三个步骤，即数据挖掘预处理，数据挖掘，数据挖掘后处理。

**7-14 有哪几种比较常用的知识发现方法？试略加介绍。**

答：常用的知识发现方法有

(1) 统计方法

统计方法是从事物外在数量上的表现去推断事物可能的规律性，包括传统方法，模糊集，支持向量机，粗糙集

 (2) 机器学习方法

包括规则归纳，决策树，范例推理，贝叶斯信念网络，科学发现，遗传算法

(3) 神经计算方法

常用的有多层感知器，反向传播网络，自适应映射网络

 (4) 可视化方法

使用有效的可视化界面，可以快速，高效地与大量数据打交道，以发现其中隐藏的特征，关系，模式和趋势

**7-15 知识发现的应用领域有哪些？试展望知识发现的发展和应用前景。**

答： (1) 金融业

数据清理，金融市场分析和预测，账户分类，银行担保和信用评估

(2) 保险业

通过对索赔者的资料与索赔历史数据模式进行比较，以判定用户的索赔是否合理

 (3) 制造业

零部件故障诊断，资源优化，生产过程分析

(4) 市场和零售业

销售预测，库存需求，零售点选择和价格分析

 (5) 医疗业

数据清理，预测医疗保健费用

 (6) 司法

案件调查，诈骗检测，洗钱认证，犯罪组织分析

(7) 工程与科学

工程与科学数据分析