

机器学习的历史与 发展

韩雁泽



目录

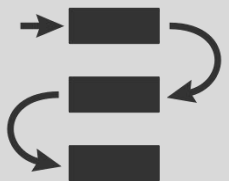
CONTENTS

01. 机器学习的定义
02. 机器学习的发展历史
03. 机器学习分类
04. 机器学习的应用场合
05. 机器学习的未来发展与挑战

01 机器学习的定义



机器学习的定义



机器学习（Machine Learning）本质上就是让计算机自己在数据中学习规律，并根据所得到的规律对未来数据进行预测。



机器学习包括如聚类、分类、决策树、贝叶斯、神经网络、深度学习（Deep Learning）等算法。



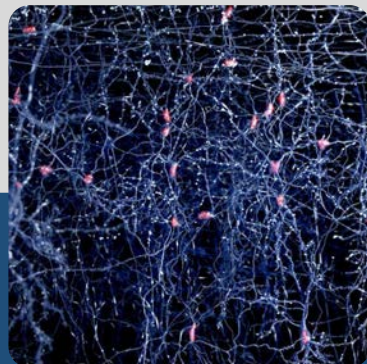
02 机器学习的发展历史

机器学习的发展历史



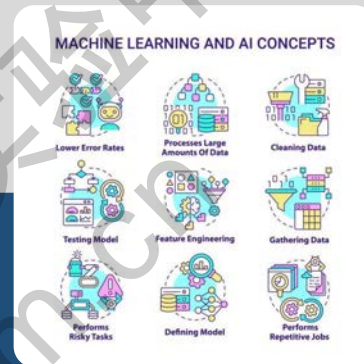
机器学习的起源

20世纪50年代，人工智能的概念被提出，旨在通过计算机模拟人类的智能行为。



神经网络的探索

20世纪80年代，神经网络（BP）的研究开始兴起，旨在通过模拟人脑神经元网络来处理信息。



支持向量机等算法的出现

随着机器学习的发展，支持向量机、朴素贝叶斯等算法相继出现，将机器学习从知识驱动转变为数据驱动的思路。



深度学习火热

2012年，随着算力提升和海量训练样本的支持，深度学习成为机器学习研究热点，并带动了产业界的广泛应用。

03 机器学习分类

监督学习 (Supervised Learning)

定义

监督学习是从已有的训练数据集（标记数据）中学习模型，然后对新的数据进行预测的一种机器学习方法。

常见算法

线性回归：通过找到最佳拟合线来预测因变量的值。

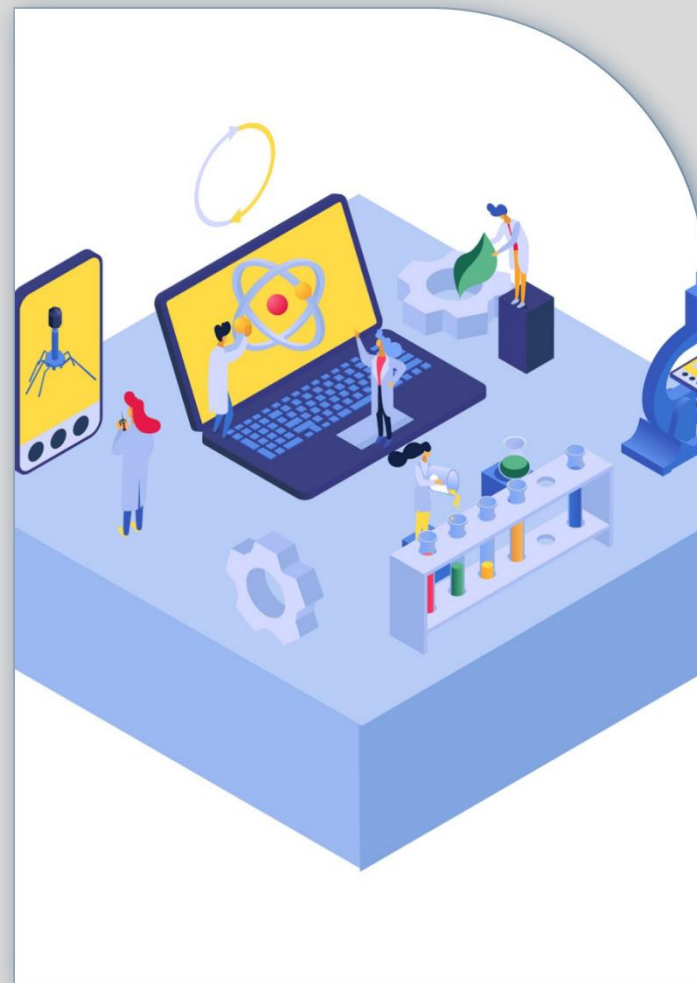
逻辑回归：用于二元分类的监督学习算法

决策树：通过构建树状结构来对新的数据进行分类或回归。

应用场景

分类问题：垃圾邮件识别、人脸识别等。

回归问题：股票价格预测、预测天气等。



无监督学习 (Unsupervised Learning)

定义

通过分析输入数据的特点和结构，自动地找出数据中的模式和规律，而不需要人工标注和干预。

常见算法

K-means：用于聚类分析。

DBSCAN：基于密度的算法，发现任意形状的聚类。

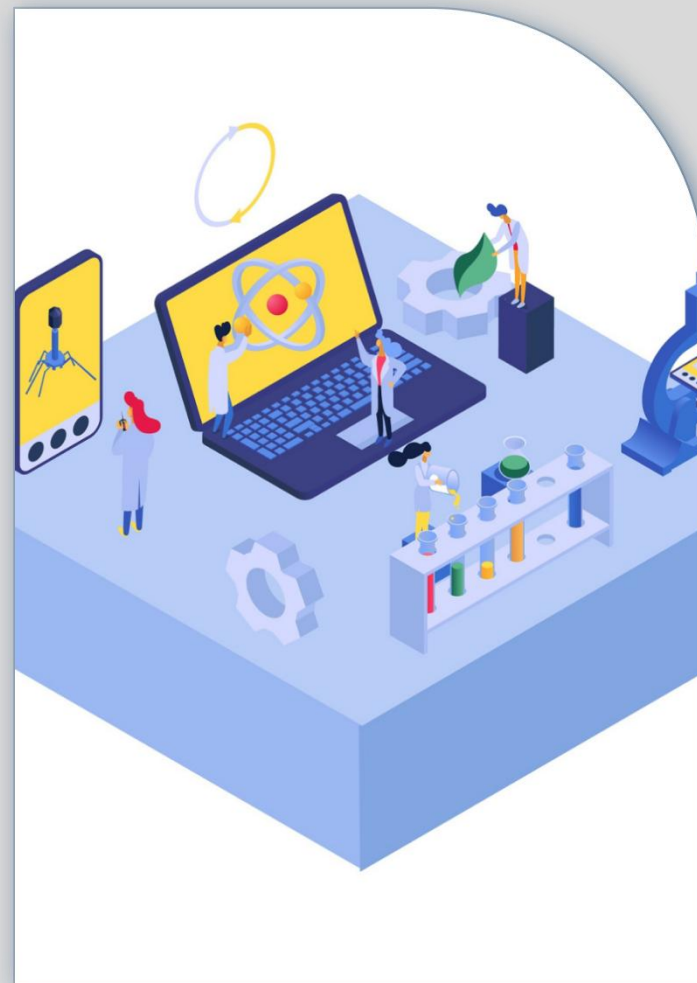
层次聚类：基于距离的算法，将数据点按照距离远近进行聚类。

应用场景

聚类：将数据集划分为多个组。

降维：将高维数据降维，更容易理解和可视化数据。

关联规则学习：超市购物篮分析中，发现哪些商品经常一起被购买。





半监督学习 (Semi-supervised Learning)

定义

利用标记和未标记的数据来进行训练和预测。

常见算法

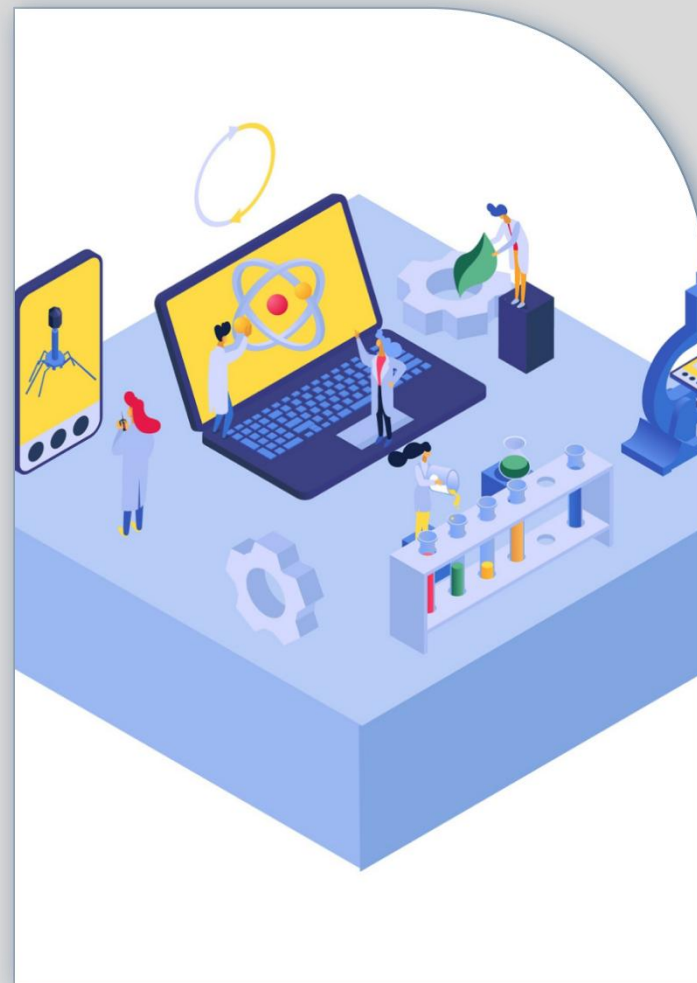
标签传播：通过迭代地传播标签，使得每个样本的标签都尽可能地一致。

学习算法：通过学习算法可以训练出更好的模型，提高分类准确率。

应用场景

分类问题：垃圾邮件识别、人脸识别等。

聚类问题：市场细分、社交网络分析等。



强化学习 (Reinforcement Learning)

定义

通过试错的方式让机器学习如何做出最优决策。

常见算法

Q-Learning: 构建Q表来对环境进行建模实现决策。

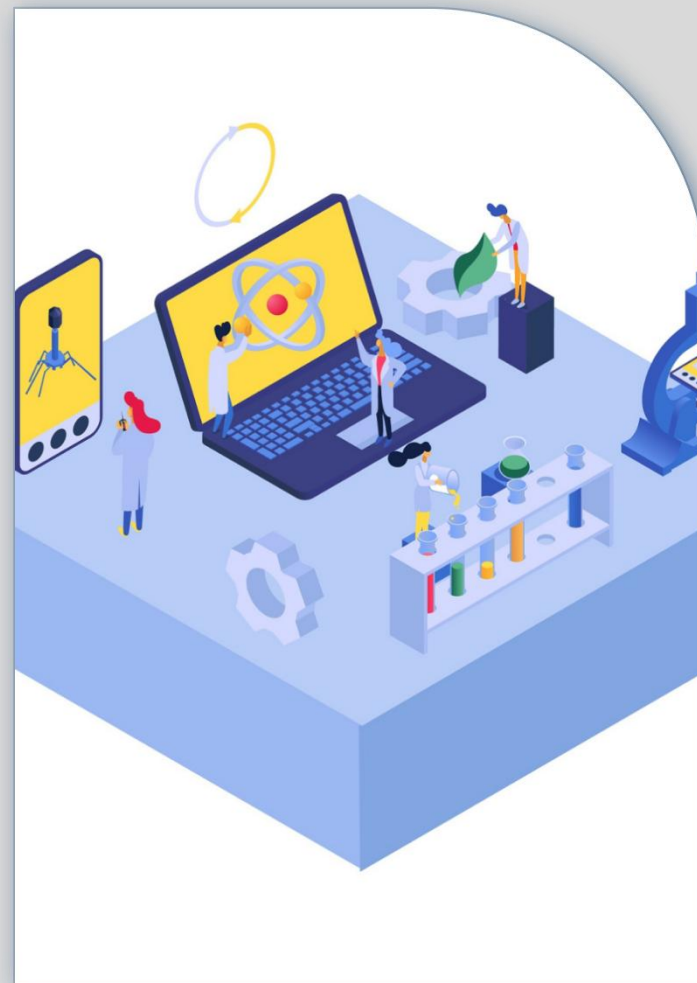
Deep Q Network (DQN): 结合深度学习通过训练神经网络来逼近Q函数, 实现更高效的学习。

Policy Gradient Methods: 优化策略寻找最优解。

应用场景

游戏AI: AlphaGo, 通过自我对弈提升技能。

机器人控制: 根据环境反馈进行自我调整, 实现更精准的控制。



04 机器学习的应用 场合



01

机器翻译

机器学习在自然语言处理领域的一个重要应用是机器翻译，它能够自动将一种语言的文本转换为另一种语言的文本。

02

语音识别

通过机器学习技术，我们现在能够使用语音助手来控制智能家居设备、查询信息或执行其他任务。

03

文本分类

机器学习算法能够对大量文本数据进行自动分类，例如垃圾邮件过滤器就是利用这一技术来识别垃圾邮件。



01

视觉垂直领域

计算机视觉在多个领域都有应用，如安防、医疗、交通等

02

预测与推荐系统

机器学习在预测和推荐系统中也有广泛的应用，如销售预测、个性化推荐等。

03

决策支持与智能分析

助分析大量数据，辅助决策制定。基于数据的决策可以更加准确和有据可依。

05

机器学习的未来 发展与挑战

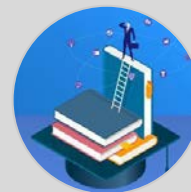


很多传统的垂直领域都需要有机器学习的加入，例如医疗、市场等，这些垂直领域可以是专门完成具体的某一项任务的模型，也可以是一个什么都会的模型。



垂直性（落地化）

机器学习衍生出来的算法、模型多种多样，它们都各有所长，也有所短，受到现场、部署环境、硬件条件、数据等影响，多样性的发展可以适配不同的领域。



多样性



数据质量问题

数据量不足

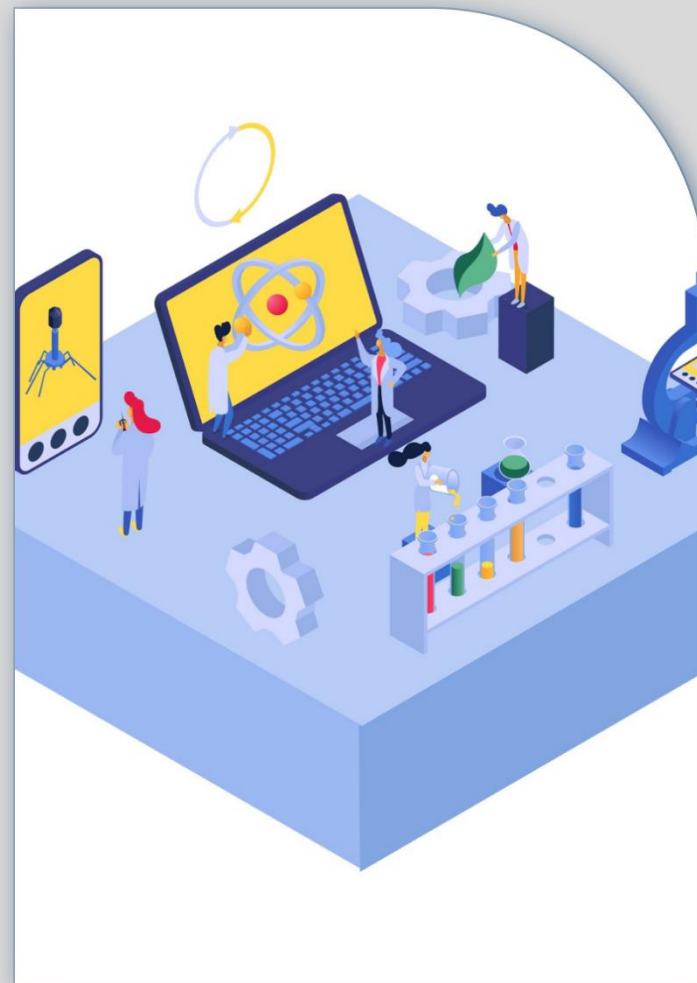
随着机器学习应用的普及，数据量不足的问题愈发突出，影响模型的训练效果和预测精度。

数据质量不均

不同来源的数据可能存在质量差异，导致模型在某些方面的性能受到影响。

数据隐私保护

在利用数据进行机器学习时，如何保护个人隐私和数据安全成为亟待解决的问题。





人工智能与安全问题

安全

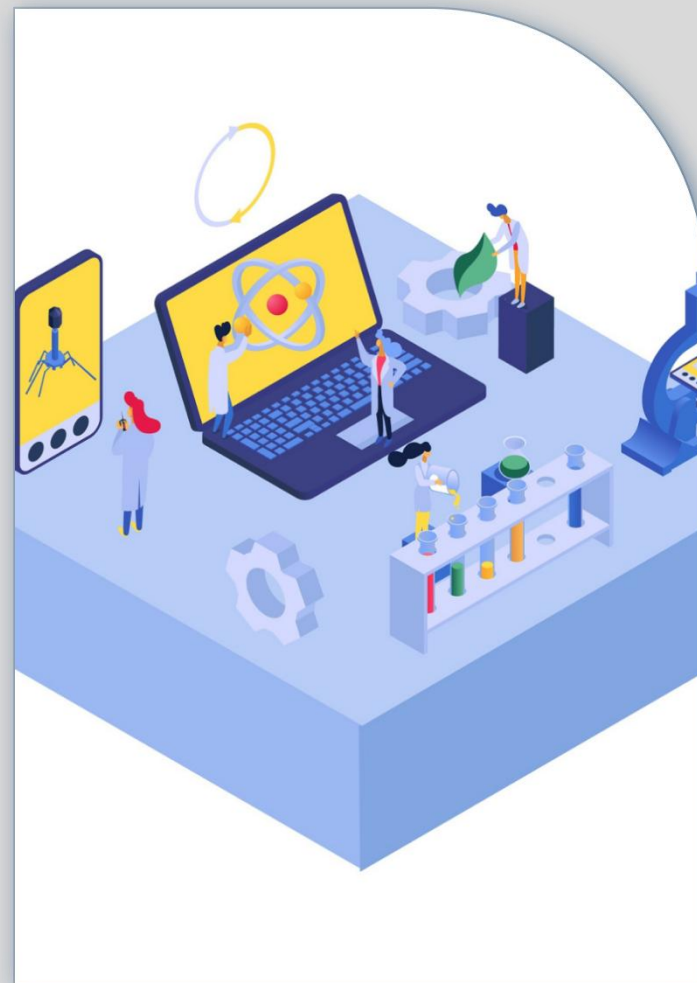
随着人工智能技术的不断发展，我们面临着越来越多的伦理挑战，例如隐私、安全和公平等问题。

人工智能的未来

需要思考如何平衡人工智能技术的发展和伦理、安全问题，以确保人工智能技术能够为人类带来更多的利益。

不可解释

随着机器学习模型变得越来越复杂，算法的可解释性变得越来越困难。



科技赋能 · 智引未来



技术领先
品质保障



超多干货
实时更新



海量视频
贴身学习

了解更多信息请访问：

集团官网：www.hqyj.com

创客学院网址（在线课程）：www.makeru.com.cn

研发中心网址（研发产品）：www.fsdev.com.cn

元宇宙实验中心网址（在线实验）：www.yyzlab.com.cn