

1 大家好，我是北京理工大学计算机学院数据科学与知识工程研究所的车海莺，本节我们讨论大数据基本概念

2 这个主题是由 Clive Humby 在 2006 年 Kellogg School 的全日空高级市场峰会上与他的公司 DunnHumby 一起率先使用数据的人提出的。

Gerd Leonhard 于 2009 年再次提出这一主题，并已成为其一部分主流

Mark Getty 是 Getty 家族的一员，该家族将其石油帝国出售给了德士古。

他想找点事情做，并决定有机会巩固摄影市场。在 2000 年《经济学人》的一篇文章 (<http://www.economist.com/node/288515>) 中，

他说：“智能资产，”他说，“是 21 世纪的石油。

看看一百年前的最富有的人：他们都靠开采自然资源或搬运自然资源赚钱。今天所有最富有的人都是从智能资产中赚钱的。”

3 我们说数据就像原油……，因为通常情况下，原始数据不能直接使用，必须经过处理才能真正应用。像原油一样需要精炼后才能使用。

大量的数据通常有洞察力，但不容易找到，数据科学家需要分析数据并找到其背后的模式。

一个数据集可以适用于多种不同的产品。这意味着一个数据集可以用于不同的目的。

比如淘宝交易数据可以用来分析不同产品的销量，找出哪些产品销量好，哪些产品销量不好。

还可以分析相同的数据以找到不同客户的购买偏好，并向他们推荐定制的相关产品。

4 生成/消费数据的模型已经改变，在旧模型中，很少量的公司在生成数据，所有其他公司都在消费数据。报社、出版社、广播电视台等少数企业是数据的来源。

在新模式中，每个人都可以成为数据的资源，我们都在生成数据，我们都在消费数据。

我们在博客、Twitter、Facebook、微信和 YouTube、TikTok 抖音等上发布内容，我们也在那里阅读内容。

5 现在来了，什么是大数据，是不是数据变大了就叫大数据？答案是否定的。

大数据是一个描述大量、难以管理的数据——结构化和非结构化数据。

这不仅体积大，而且我们以前使用的传统方式也很难管理。

我们过去处理结构化数据的传统方式，即具有行和列的二维表形式的数据。

—日复一日的数据正在淹没企业的业务数据。

...重要的是组织如何处理数据。可以分析大数据以获得洞察力，从而做出更好的决策和战略性业务举措。

6 从电子数据的产生到今天，我们经历了几个阶段。早期，交易系统随着业务活动的发生而产生的数据。比如你在沃尔玛买东西，它会在沃尔玛系统中生成一个销售记录数据。

然后随着社交媒体和移动设备的发展，普通人可以发送消息并以文本语音和视频的形式发布内容。所以这是用户生成的内容阶段。

第三阶段——感知阶段具有处理能力的传感器，自动；(7*24) 生成数据，像摄像机、可穿戴设备等，一直在生成数据，没有停止。

---这是大数据的主要原因

7 这张图展示了 IT 的进步，自 1980 年代冯诺依曼发明计算机以来，我们进入了信息时代，PC 可以生成和处理数据。然后互联网将所有计算机连接在一起，使数据在所有 PC 之间移动和共享，从而促进了信息的生成和交换。信息可以很容易地从一个地方传输到地球的另一端。移动设备的普及，解放了电脑对人们

物理空间的限制，人们可以随时随地创作和发布内容。这使得大数据。纽约时报 2012 专栏：大数据时代已经到来，决策更依赖数据分析。我们有很多来自不同方面的数据，连接所有这些数据并以合适的方式处理它，我们可以提取有用的信息来支持我们的决策。2010 年之后，人工智能时代已经来临，成为继互联网之后的第二次科技生态萌芽。AI 赋予机器一些智能，可以分析数据本身，帮助人们分析数据并做出相应的反应。这使我们能够更广泛、更深入地分析，从而获得更多洞察力。四大技术赋能大数据、人工智能、物联网、5G

8 在本节中，我们学习了大数据的概念，IT 发展的历程以及大数据是如何产生的，感谢您的学习，如果您有任何问题，请随时与我联系。