

## 预习 03 | 大数据应用领域：数据驱动一切

2018-11-03 李智慧

从0开始学大数据

[进入课程 >](#)



讲述：李智慧

时长 11:21 大小 5.21M



大数据出现的时间只有十几年，被人们广泛接受并应用只有几年的时间，但就是这短短几年的时间，大数据呈现出爆炸式增长的态势。在各个领域，大数据的身影几乎无处不在。今天我们通过一些[大数据典型的应用场景分析](#)，一起来看看大数据到底能做些什么，我们学大数据究竟有什么用，应该关注大数据的哪些方面。

### 大数据在医疗健康领域的应用

健康医疗领域是最近几年获得最多创业者和投资人青睐的大数据领域。为什么这么说呢？首先，医疗健康领域会产生大量的数据；其次，医疗健康领域有一个万亿级的市场规模；最关键的是，**医疗健康领域里很多工作依赖人的经验，而这正是机器学习的强项。**

#### 1. 医学影像智能识别

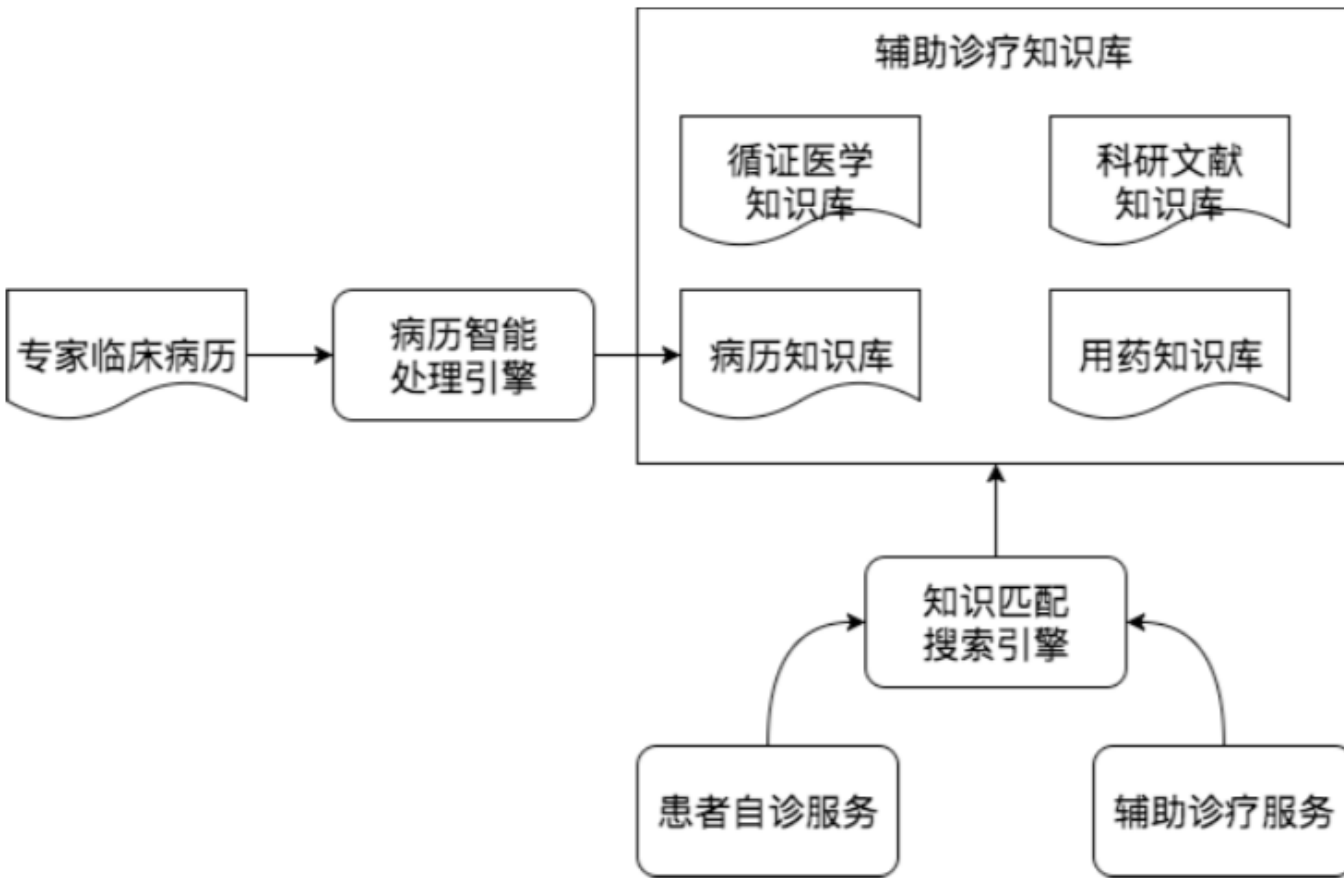
图像识别是机器学习获得的重大突破之一，使用大量的图片数据进行深度机器学习训练，机器可以识别出特定的图像元素，比如猫或者人脸，当然也可以识别出病理特征。

比如 X 光片里的异常病灶位置，是可以通过机器学习智能识别出来的。甚至可以说医学影像智能识别在某些方面已经比一般医生拥有更高的读图和识别能力，但是鉴于医疗的严肃性，现在还很少有临床方面的实践。

虽然在临床实践方面应用有限，但是医疗影像 AI 还是在一些领域取得一定的进展。医学影像智能识别，一方面可以帮助医生进行辅助诊疗，另一方面对于皮肤病等有外部表现的病症，病人可以自己拍照然后使用 AI 智能识别做一个初步诊断。

## 2. 病历大数据智能诊疗

病历，特别是专家写的病历，本身就是一笔巨大的知识财富，利用大数据技术将这些知识进行处理、分析、统计、挖掘，可以构成一个病历知识库，可以分享给更多人，即构成一个智能辅助诊疗系统。下面这张图是我曾经参与设计过的一个医疗辅助诊疗系统的架构。



针对同类疾病和其他上下文信息（化验结果、病史、年龄性别、病人回访信息等）可以挖掘出针对同样的疾病情况，哪种治疗手段可以用更低的治疗成本、更少的病人痛苦，获得更好

的治疗效果。从上面的架构图你能看到，将这些病历知识和循证医学知识、科研文献知识、用药知识共同构成一个辅助诊疗知识库，通过知识匹配搜索引擎可以对外提供服务。患者或者医生录入病史、检查结果等信息，系统匹配初步诊断结果，搜索诊疗计划，产生多个辅助诊疗建议，供患者和医生进行参考。

## 大数据在教育领域的应用

教育倡导“因人施教”，但是在传统教育过程中要做到因人施教，需要老师本身能力很强才能把握好。但是大数据在线教育利用大数据技术进行分析统计，完全可以做到根据学生能力和学习节奏，及时调整学习大纲和学习进度，提供个性化和自适应的学习体验。除此之外，人工智能在教育的其他方面也取得很好的进展。

### 1. AI 外语老师

得益于语音识别和语音合成技术的成熟（语音识别与合成技术同样是利用大数据技术进行机器学习与训练），一些在线教育网站尝试用人工智能外语老师进行外语教学。这里面的原理其实并不复杂，聊天机器人技术已经普遍应用，只要将学习的知识点设计进聊天的过程中，就可以实现一个简单的 AI 外语老师了。

### 2. 智能解题

比较简单的智能解题系统其实是利用搜索引擎技术，在收集大量的试题以及答案的基础上，进行试题匹配，将匹配成功的答案返回。这个过程看起来就像智能做题一样，表面看给个题目就能解出答案，而实际上只是找到答案。

进阶一点的智能解题系统，通过图像识别与自然语言处理（这两项技术依然使用大数据技术实现），进行相似性匹配。更改试题的部分数字、文字表述，但是不影响实质性解答思路，依然可以解答。

高阶的智能解题系统，利用神经网络机器学习技术，将试题的自然语言描述转化成形式语言，然后分析知识点和解题策略，进行自动推导，从而完成实质性的解题。

## 大数据在社交媒体领域的应用

大数据有一个重要的、和我们大多数人密切相关，但是又不太引人注目的一个应用领域是**舆情监控与分析**。我们日常在各种互联网应用和社交媒体上发表各种言论，这些言论事实上反

映了最准确的民情舆论。一个个体的言论基本没有意义，但是大量的、全国乃至全球的言论数据表现出的统计特性，就有了非常重要的意义。

编写数据爬虫，实时爬取各个社交新媒体上的各种用户内容和媒体信息，然后通过自然语言处理，就可以进行情感分析、热点事件追踪等。舆情实时监控可用于商业领域，引导智能广告投放；可用于金融领域，辅助执行自动化股票、期权、数字货币交易；可用于社会管理，及时发现可能引发社会问题的舆论倾向。

在美国总统大选期间，候选人就曾雇佣大数据公司利用社交媒体的数据进行分析，发现选票可能摇摆的地区，有针对性前去进行竞选演讲。并利用大数据分析选民关注的话题，包装自己的竞选主张。Facebook 也因为授权大数据公司滥用自己用户的数据而遭到调查和谴责，市值蒸发了数百亿美元。

## 大数据在金融领域的应用

大数据在金融领域应用比较成熟的是**大数据风控**。在金融借贷中，如何识别出高风险用户，要求其提供更多抵押、支付更高利息、调整更低的额度，甚至拒绝贷款，从而降低金融机构的风险？事实上，金融行业已经沉淀了大量的历史数据，利用这些数据进行计算，可以得到用户特征和风险指数的曲线（即风控模型）。当新用户申请贷款的时候，将该用户特征带入曲线进行计算，就可以得到该用户的风险指数，进而自动给出该用户的贷款策略。

利用股票、外汇等历史交易记录，分析交易规律，结合当前的新闻热点、舆论倾向、财经数据构建交易模型，进行自动化交易，这就是金融领域的量化交易。这些数据量特别巨大，交易涉及金额也同样巨大，所以金融机构在大数据领域常常不惜血本，大手笔投入。

## 大数据在新零售领域的应用

区别于传统零售，新零售使用大数据进行全链路管理。从生产、物流、购物体验，使用大数据进行分析和预判，实现精准生产、零库存、全新的购物体验。

亚马逊 Go 无人店使用大量的摄像头，实时捕捉用户行为，判断用户取出还是放回商品、取了何种商品等。这实际上是大数据流计算与机器学习的结合，最终实现的购物效果是，无需排队买单，进去就拿东西，拿好了就走，超级科幻有没有。

虽然无人店现在看起来噱头的意味更多一点，但是利用大数据技术提升购物体验、节省商家人力成本一定是正确的方向。

## 大数据在交通领域的应用

交通也是一个对大数据实时采集与处理应用比较广的领域。现在几乎所有的城市路段、交通要点都有不止一个监控摄像头在实时监控，一线城市大约有百万计的摄像头在不停地采集数据。这些数据一方面可以用于公共安全，比如近年来一些警匪片里会有一些场景：犯罪嫌疑人驾车出逃，警方只要定位了车辆，不管它到哪里，系统都可以自动调出相应的摄像头，实时看到现场画面。应该说这项技术已经成熟，大数据流计算可以对百万计的流数据实时处理计算，电影里的场景计算其实并不复杂。

此外，各种导航软件也在不停采集数据，通过分析用户当前位置和移动速度，判断道路拥堵状态，并实时修改推荐的导航路径。你如果经常开车或者打车，对这些技术一定深有体会。

还有就是无人驾驶技术，无人驾驶就是在人的驾驶过程中实时采集车辆周边数据和驾驶控制信息，然后通过机器学习，获得周边信息与驾驶方式的对应关系（自动驾驶模型）。然后将这个模型应用到无人驾驶汽车上，传感器获得车辆周边数据后，就可以通过自动驾驶模型计算出车辆控制信息（转向、刹车等）。计算自动驾驶模型需要大量的数据，所以我们看到，这些无人驾驶创业公司都在不断攀比自己的训练数据有几十万公里、几百万公里，因为训练数据的量意味着模型的完善程度。

## 小结

正如我前面所说，利用大数据和机器学习，发掘数据中的规律，进而对当前的事情做出预测和判断，使机器表现出智能的特性，正变得越来越普及。

大数据主要来自企业自身所产生，还有一些数据来自互联网，通过网络爬虫可以获取；再有就是公共数据，比如气象数据等。所有这些数据汇聚在一起，计算其内在的关系，可以发现很多肉眼和思维无法得到的知识。然后进一步计算其内在的模型，可以使系统获得智能的特性。当系统具备智能的特性，可以使机器对当前的事情做出预测和判断，正如我今天和你聊的，大数据技术应用正变得越来越普及。

但是，这些数据通常非常巨大，存储、计算、应用都需要一套不同以往的技术方案。通过前面 3 期内容，我带你了解了大数据技术和应用的发展史，以及当今大数据典型的应用领域。从第 4 期开始，我将会从大数据主要产品的架构原理、大数据分析与应用、数据挖掘与机器学习算法等几个维度，全面讲解大数据的方方面面，相信你一定有所收获。

## 思考题

除了我今天聊的这几个领域应用了大数据技术，你还能举出其他典型的大数据应用吗？你知道它实现的原理吗？

欢迎你写下自己的思考或疑问，与我和其他同学一起讨论。



# 从 0 开始学大数据

## 智能时代你的大数据第一课

**李智慧**  
同程艺龙交通首席架构师  
前 Intel 大数据架构师



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 预习 02 | 大数据应用发展史：从搜索引擎到人工智能

下一篇 04 | 移动计算比移动数据更划算

## 精选留言 (57)

 写留言



凯恩-格吉...

2018-11-03

 96

一家快餐披萨店，外卖电话响了，店长拿起电话。

店长：您好，这里是××披萨店。请问有什么需要我为您服务？

顾客：你好，我想要订一份披萨。

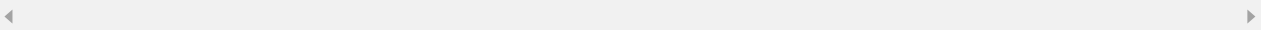
店长：请问您是陈先生吗？

顾客：你怎么知道我姓陈？ ...



展开 ▾

作者回复: ㊦



**My dream**

2018-11-03

👍 31

老师，我建议你在课程中讲一下实战操作分析数据的步骤，让我们理解的更深刻直观



**Hyun**

2018-11-03

👍 21

首先得问自己一个问题，即我能说出的应用领域有多少呢？

医疗健康

餐饮、食品、农业

交通运输...

展开 ▾



**余生、请你...**

2018-11-05

👍 12

按照 凯恩-格吉塔斯 网友这么说，人在大数据面前还有多少隐私？

展开 ▾



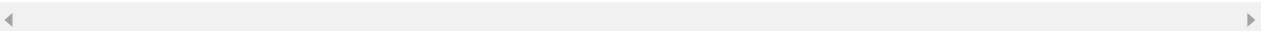
**lyshrine**

2018-11-03

👍 10

我想到的还有体育领域，可以利用大数据制定球队战略，利用实时数据及时调整比赛中的战术。还可以对体育训练有帮助，可以分析和总结优秀选手的动作姿势，纠正其它他运动员的动作。未来体育仅仅靠天赋和苦练不足以取得最好的成绩，必须依赖大数据。

作者回复: 赞



**S.H.**

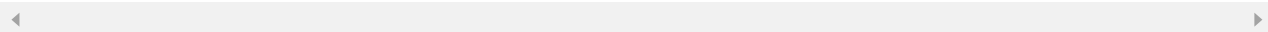
2018-11-06

👍 5

对我们这门课程学习来说，可以通过获取用户的学习时长，学习时间，页面停留时间，反复阅读次数，评论交互等数据信息，分析用户的学习需求，平台可以用来做课程推荐，讲师可以优化课程。

展开 ▾

作者回复: 是的



**张相荣**

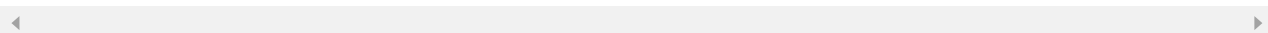
2018-11-03

👍 5

现代化智慧农业也是大数据技术服务的一个场景

展开 ▾

作者回复: 是的



——

2018-11-03

👍 5

我想到两点，1 公交车通过大数据技术分析计算每辆公交车每天的用电或油量，不断优化驾驶员驾驶习惯，从而达到节省城市公交成本的目的。2 外卖平台通过大数据技术实时定位每位外卖员的位置和其手头的接单数量，不断优化接单速度和订单完成速度。

作者回复:

想法不错👍



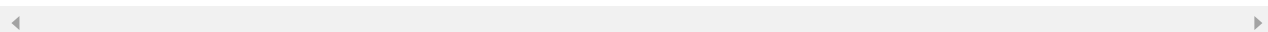
**Austin**

2018-11-13

👍 4

城市摄像头收集行为数据，可以根据行为，比如说站姿，走路方式，等人类无法发现的细节，对指定人进行追踪，无论他是否露脸，永远逃不掉。

作者回复: 厉害







Cs

2018-12-14

👍 2

我更觉得大数据分析是对整个群体而言，个体意义对于这个没太大的意义，也就不是担心个人隐私会泄露的问题.....即使是应对措施也是针对群体而言的，像前面说的那样，能准确判断出个人的一切信息以及决策，那本质上已经侵犯了个人权益，这应该还是要避免的



李渝Will

2018-12-12

👍 2

大数据的应用太多了 关键是用之后有什么质的效果 不能画蛇添足 为了数据分析而分析



小千

2018-11-05

👍 2

其实目前公认的大数据技术优于传统技术的也就是三个方向:

- 1.图像，语音识别，生成。
- 2.推荐，搜索系统。
- 3.简单规则的游戏。

大数据能成功应用的领域，都是把这些领域的问题成功的转换成上述三个技术问题，比...  
展开 ▾



落叶飞逝的...

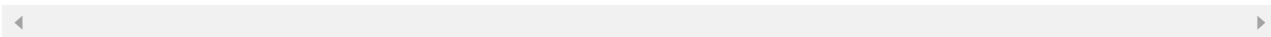
2018-11-05

👍 2

购物猜你喜欢，新闻、段子都是大数据的应用

展开 ▾

作者回复: 是的



ming

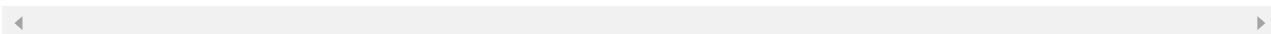
2018-11-05

👍 2

抖音、头条的智能推荐也是基于大数据

展开 ▾

作者回复: 是的，BAT在PC时代崛起，美团在移动时代崛起，头条在智能时代崛起，头条抓住了时代的潮流。





**Backkom**

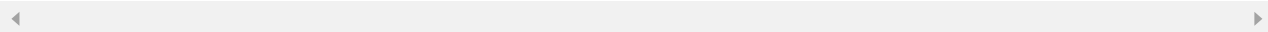
2018-11-04

👍 2

天气预报，海量天气因素数据采集，结合同期过往的大量数据，进行预测分析未来N天的天气走势。实时采集实时计算可以显示当前级未来N小时的天气情况。

展开 ▾

作者回复: 赞



**无聊才读书**

2018-12-01

👍 1

但是感觉没有隐私了！

展开 ▾



**张苗**

2018-11-03

👍 1

我们每天都会接触到的电商app，新闻app中的商品推荐，资讯推荐功能都会用到大数据技术。后台系统会统计我们的购买习惯，资讯偏好来给我们推荐想要的东西。

展开 ▾

作者回复: 是的



**探索无止境**

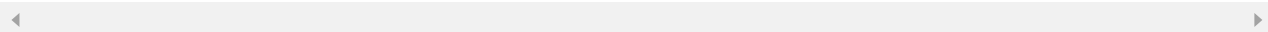
2018-11-03

👍 1

大数据还可以结合人脸识别，帮助警方尽快定位到嫌疑人，加快破案速度

展开 ▾

作者回复: 是的



**李铁牛**

👍 1

2018-11-03

我觉得环境方面应用大数据也比较广，像是天气预报，空气质量预测等等，而且还可以和很多行业结合，比如交通，医疗

作者回复: 是的



**lele\_offer**

2019-04-22



能产生大量数据的领域，  
依赖人经验的领域，  
场景中多变，人类预测困难的领域，  
提高预测速度，准确度能带来价值的领域，  
在人性的优点和弱点之间适当平衡的领域， ...

展开 ∨